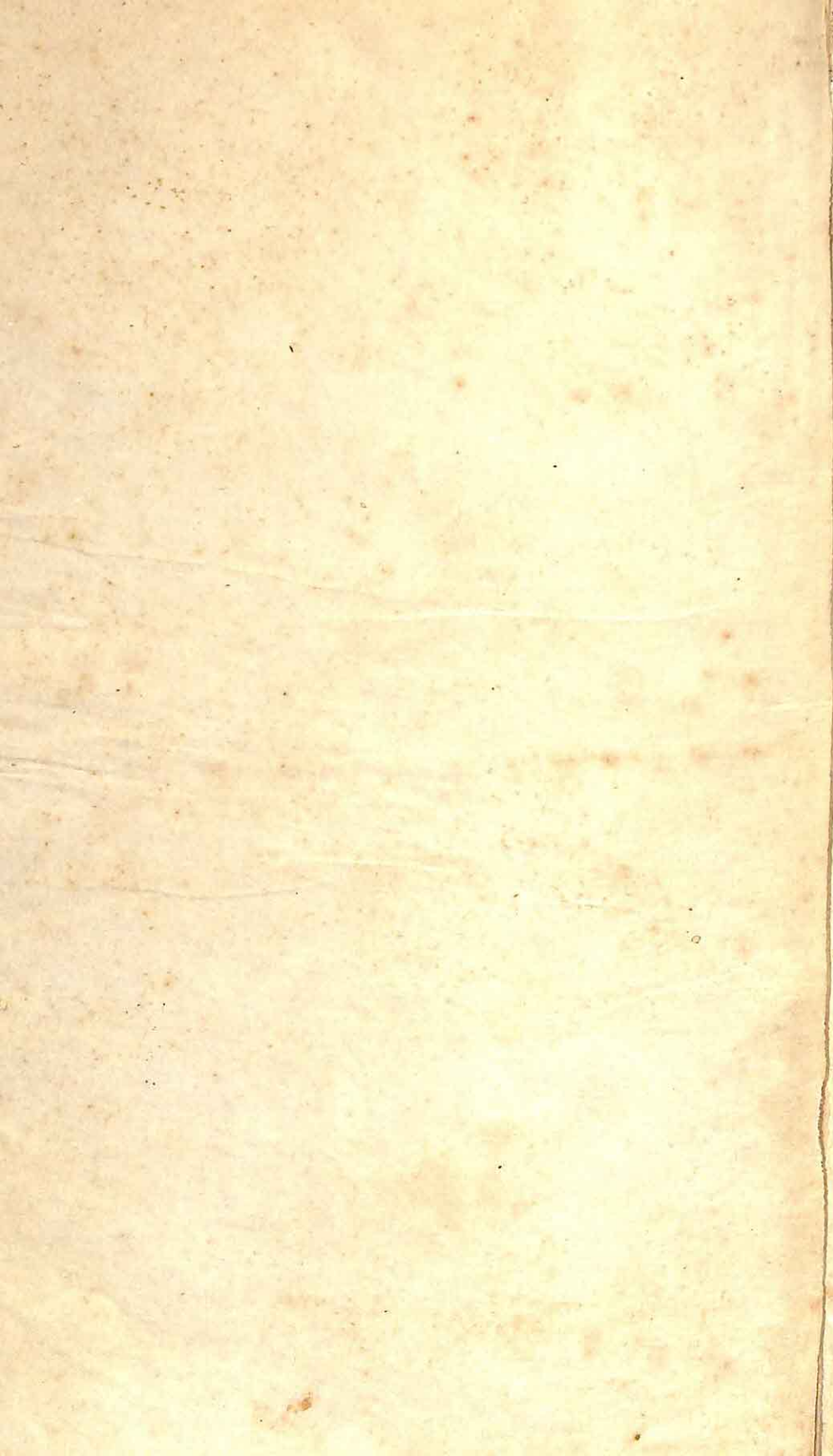


145
12.8.76



Psychological Bulletin

EDITED BY

LYLE H. LANIER, VASSAR COLLEGE

478 Z 89

WITH THE CO-OPERATION OF

S. H. BRITT, McCANN-ERICKSON, INC., NEW YORK; D. A. GRANT, UNIVERSITY OF WISCONSIN; W. T. HERON, UNIVERSITY OF MINNESOTA; W. A. HUNT, NORTH-WESTERN UNIVERSITY; J. G. JENKINS, UNIVERSITY OF MARYLAND; D. G. MARQUIS, UNIVERSITY OF MICHIGAN; A. W. MELTON, OHIO STATE UNIVERSITY; J. T. METCALF, UNIVERSITY OF VERMONT.

VOLUME 44, 1947



PUBLISHED BI-MONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1515 MASSACHUSETTS AVE., N.W.,
WASHINGTON 5, D. C.

Edm. Research
12.5.70
9 145

CONTENTS OF VOLUME 44

ORIGINAL CONTRIBUTIONS, GENERAL REVIEWS, DISCUSSIONS

The "laws" of relative variability of mental traits: ROBERT S. ELLIS.	1
A discussion of some causes of operational fatigue in the Army Air Forces: LESSING A. KAHN.	34
A review of leadership studies with particular reference to military problems: WILLIAM O. JENKINS.	54
Comment on "The validity of personality questionnaires": CHRISTIAN PAUL HEINLEIN.	80
Discussion of Heinlein's comment on "The validity of personality questionnaires": ALBERT ELLIS.	83
The masking of speech: GEORGE A. MILLER.	105
The effective use of manipulative tests in industry: W. F. LONG AND C. H. LAWSHE, JR.	130
The treatment of qualitative data by "scale analysis": LEON FESTINGER.	149
Manual of child psychology: ROGER G. BARKER.	162
Response to Crespi's rejoinder and Conrad's reply to appraisal of opinion-attitude methodology: QUINN MCNEMAR.	171
Convulsive behavior in the rat: FRANK W. FINGER.	201
A critical review of the literature on "absolute pitch": D. M. NEU.	249
Selective sampling in psychological research: ELI S. MARKS.	267
Additional tables of the probability of "runs" of correct responses in learning and problem-solving: DAVID A. GRANT.	276
Encyclopedia of psychology: S. RAINS WALLACE.	280
The relation of emotional adjustment to intellectual function—a note: ROBERT E. HARRIS and CLARE WRIGHT THOMPSON.	283
Corrigenda: H. M. JOHNSON.	288
The present status of psychotherapeutic counseling : WILLIAM U. SNYDER.	297
Spontaneous activity of animals: J. DAVID REED.	393
Sampling in the revision of the Stanford-Binet scale: ELI S. MARKS	413
Illumination standards for effective and easy seeing: MILES A. TINKER.	435
On Festinger's evaluation of scale analysis: LOUIS Guttman.	451

Note on "A review of leadership problems with particular reference to military problems": DONALD E. BAIER.....	466
The history and present status of the law of effect: LEO POSTMAN..	489
Discussion of Snyder's "The present status of psychotherapeutic counseling": J. L. MORENO.....	564

BOOK REVIEWS

Masserman's Principles of dynamic psychiatry: EDWIN R. GUTHRIE.....	87
Rogers and Wallen's Counseling with returned servicemen: WILLIAM A. HUNT.....	88
Johnson's People in quandaries: IRWIN AUGUST BERG.....	89
Carp's A study of the influence of certain personal factors on a speech judgment: ST. CLAIR A. SWITZER.....	90
Morgan's How to keep a sound mind: CLAUDE M. DILLINGER....	90
Stern's The attendant's guide: STANLEY S. MARZOLF.....	91
Huntington's Mainsprings of civilization: DALE B. HARRIS.....	92
Smith, Lasswell and Casey's Propaganda, communication and public opinion; a comprehensive reference guide: R. B. AMMONS.....	94
Dewey's Problems of men: WESLEY R. WELLS.....	95
Rand, Sweeny and Vincent's Growth and development of the young child: GEORGE G. THOMPSON.....	96
Beaumont's Psychology of personnel and Psychology applied to personnel: THOMAS W. HARRELL.....	97
Maier's Psychology in industry: LESTER P. GUEST.....	99
Harriman's (Ed.) Twentieth century psychology: GEO. M. PETERSON.....	101
Symonds' The dynamics of human adjustment: FRED McKINNEY.....	177
Benedek's Insight and personality adjustment: a study of the psychological effects of war: THOMAS W. RICHARDS.....	179
Hewitt and Jenkins' Fundamental patterns of maladjustment; the dynamics of their origin: C. LANDIS.....	180
Thorpe's Child psychology and development: KATHERINE M. Maurer.....	181
Schroeder's On measurement of motor skills; an approach through a statistical analysis of archery scores: JAMES M. LYNCH.....	182
Edwards' Statistical analysis for students in psychology and education: JOSEPH ZUBIN.....	183

Gottschalk and Angell's The use of personal documents in history, anthropology and sociology: ARTHUR F. JENNESS.	185
Lillie's General biology and philosophy of organism: ROBERT P. FISCHER.	187
Fales' Wisdom and responsibility—an essay on the motivation of thought and action: GOERGE W. HARTMANN.	188
Sargent's Teach yourself psychology: LLOYD G. HUMPHREYS. ...	189
Blumenfeld's Introducción a la psicología experimental: HOWARD DAVIS SPOERL.	190
Link and Hopf's People and books: NANCY C. COOLEY AND ROBERT H. SEASHORE.	191
Cleeton and Mason's Executive ability: its discovery and development: WILLIAM O. JENKINS.	192
Blair's Diagnostic and remedial teaching in secondary schools: JACOB I. HARTSTEIN.	193
Deutsch's The mentally ill in America: a history of their care and treatment from colonial times: JULES D. HOLZBERG.	194
Dunlap's Personal adjustment: BERT R. SAPPENFIELD.	195
Baxter's An approach to guidance: ALBERT S. THOMPSON.	196
Black's Off to a good start: MIRIAM FORSTER FIEDLER.	197
Dunlap's Religion: its functions in human life: ERNEST M. LIGON.	289
Friedman's Foundations of the measurement of values: ELIZABETH DUFFY.	290
Mays' The concept of vocational education in the thinking of the general educator, 1845 to 1945 : JAMES L. GRAHAM. ...	292
Kingsley's The nature and conditions of learning: JAMES M. PORTER JR.	292
Traxler and Townsend's Another five years of research in reading: WILLIAM A. BROWNELL.	294
Moncrieff's The chemical senses: B. VON HALLER GILMER.	295
Roe and Burks' Adult adjustment of foster children of alcoholic and psychotic parentage and the influence of the foster home: ROBERT C. CHALLMAN.	387
Shakow's The nature of deterioration in schizophrenic conditions: WALTER C. SHIPLEY.	388
Thornton's Problems in abnormal behavior : FRANK A. PATTIE. ...	389
Abrahamsen's The mind and death of a genius: ALFRED L. BALDWIN.	389
Blankenship's How to conduct consumer and opinion research: the sampling survey in operation: STEUART HENDERSON BRITT	391

Minor's Experimental hypertension: CHESTER W. DARROW.....	392
Munn's Psychology: the fundamentals of human adjustment: HARRY HELSON.....	468
Bridges' Psychology normal and abnormal: GEORGE M. HAS- LERUD.....	475
Gray's Psychology in human affairs: ARTHUR G. BILLS.....	476
Luck and Hall's (Eds.) Annual review of physiology: (Vols. VII & VIII) CLIFFORD T. MORGAN	477
Barker, Wright and Gonick's Adjustment to physical handicap and illness: a survey of the social psychology of physique and disability: ROSS STAGNER.....	479
Lewis' Children of the Cumberland: WAYNE DENNIS	480
Leeper's Psychology of personality: BERT R. SAPPENFIELD.....	482
Kelley's 22 cells in Nuremberg: LEWIS M. TERMAN.....	483
Landis and Bolles' Textbook of abnormal psychology: JOHN DOLLARD.....	568
Horsley's Narco-analysis: JOSEPH ZUBIN.....	570
Lowrey's Psychiatry for social workers: H. MELTZER.....	571
Arthur's Tutoring as therapy: ADELLA C. YOUTZ.....	572
Witmer's (Ed.) Psychiatric interviews with children: FLORENCE L. GOODENOUGH.....	573
Wolff's Personality of the preschool child: ALFRED L. BALDWIN...	575
Radke's The relation of parental authority to children's behavior and attitudes: NANCY BAYLEY.....	576
Maurer's Intellectual status at maturity as a criterion for selecting items in preschool tests: QUINN MCNEMAR.....	578
Ford's The application of the Rorschach test to young children: L. JOSEPH STONE.....	579
Kaback's Vocational personalities: an application of the Ror- schach Group Method: MAX HERTZMAN.....	581
Lawton's Aging successfully: H. S. CONRAD.....	582
de Gruchy's Creative old age: H. S. CONRAD.....	582
Michotte's La perception de la causalité: FRITZ HEIDER.....	585
Delgado's La personalidad y el carácter: HOWARD DAVIS SPOERL	586
Link's The rediscovery of morals, with special reference to race and class conflict: ERNEST R. HILGARD.....	587

MISCELLANEOUS

Books and Materials Received.....	103, 199, 485, 588
Index of subjects.....	591
Index of authors.....	593

Psychological Bulletin

THE MASKING OF SPEECH*

GEORGE A. MILLER

Psycho-Acoustic Laboratory, Harvard University

The ability to hear two sounds at once is one of the most useful properties of the human ear. This ability—noted by G. S. Ohm more than a century ago—allows us to respond selectively to certain components of the total acoustic pattern and to ignore others, to hear our friend's voice in spite of a background of noise, to follow the theme and still hear the obbligato. But, remarkable as this ability is, it is not infallible, and sometimes we are unable to hear one sound because another sound interferes. The selective mechanism is studied in its simplest form when two pure tones are introduced into the ear and the listener is asked to report the presence or absence of one or the other of the tones. In such experiments it is found that the ear is not a perfect analyzer, for some tones obscure the perception of others. This interference is called auditory masking.

Auditory masking is usually defined as "the shift of the threshold of audibility of the masked sound due to the presence of the masking sound" (23). Following this definition, the measurement of auditory masking is a straightforward experimental procedure. First, the just-audible sound-pressure is determined in the quiet. Then the interfering sound is introduced and the listener's threshold is again determined. The difference, in decibels, between the quiet and the masked thresholds is the measure of the amount of masking produced by that particular type and intensity of interference for that particular masked sound. The notion can be applied to complex noises like speech just as readily as it is applied to pure tones. The introduction of an interfering noise makes it necessary to raise the intensity of the speech for

* Much of the research reviewed here was begun under an OSRD contract and is continuing under contract with the U. S. Navy, Office of Naval Research (Contract N5ori-76, Report PNR-23).

All of the research reviewed has been published previously but because of military classification some of the publications are available to only a limited psychological audience. In some cases it has been impossible even to refer to the original source.

Vol 44 No 2 January 1947

it to be understood. The necessary increment is taken as the measure of the masking produced by the noise.

Knowledge of the ear's susceptibility to interference is obviously a matter of considerable practical value. Much of our present knowledge has grown out of the job of developing the telephone, although research on speech-communication problems in World War II has supplemented the earlier investigations. In the course of this research many different sounds have been studied to determine the interference they might produce, and a review of the results shows the masking of speech to depend on three characteristics of the masking sound: (1) its intensity relative to the intensity of the speech, (2) its acoustic spectrum, and (3) its temporal continuity. In the following pages we will review a variety of sounds which are, or might be, encountered, and in every case the disruption of vocal communication is determined by these three attributes. Human speech is most seriously masked by an uninterrupted noise which has its power concentrated in the lower third of a spectrum covering the frequency-range from 100 to 4000 or 5000 cycles.

THE SPECTRUM OF SPEECH

A noise which interferes with our perception of one sound may not interfere with our perception of another. Consequently, it is of some importance to examine the spectrum of human speech, since speech is the sound of direct concern for the present discussion.

Physically, speech consists of vibrations varying widely and rapidly in their intensity and frequency (2, 4, 19). In general, the vowels can be analyzed into discrete component frequencies, with fundamentals in the neighborhood of 100–200 cycles per second. Some consonants, however, contain energy distributed almost continuously through the high frequencies. Thus the energy is constantly shifting from one range of frequencies to another as the talker proceeds from one sound to the next.

One attempt to describe the energy in speech as a function of frequency borrows the concept of the spectrum from optics (2). A typical "long-interval speech spectrum" is shown in Fig. 1 (17). A crew of seven young men spoke an English sentence in a "conversational" voice, and a condenser microphone located 18 inches in front of the lips picked up the sound-waves and converted them into voltages which were analyzed by an audio-spectrometer. At a distance of 18 inches, the total pressure integrated over the entire range of frequencies was about 76 db above the standard reference pressure of 0.0002 dyne/cm². The spectrum shown in Fig. 1 represents the root-mean-

square pressure in frequency-bands one cycle wide. For purposes of orientation the minimum audible field for pure tones is indicated at the bottom of the graph (18). This type of representation shows the long-time average distribution of speech-energy over the range of audible frequencies, but it is necessary to remember that the "instantaneous" spectrum is constantly shifting* and is seldom, if ever, similar to the long-time average.

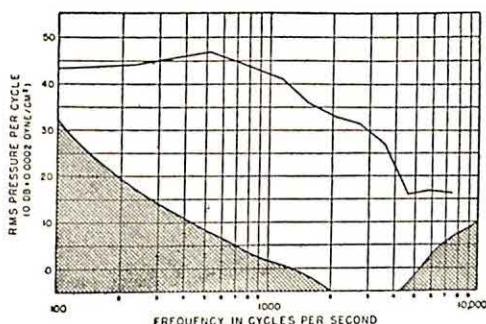


FIG. 1. THE LONG-INTERVAL SPECTRUM OF CONVERSATIONAL AMERICAN SPEECH FOR SEVEN MALE VOICES, EXPRESSED IN TERMS OF THE RMS PRESSURE IN FREQUENCY-BANDS ONE CYCLE WIDE.

The over-all level of the speech 18 inches in front of the talker's lips was 76 db re 0.0002 dyne/cm².

In terms of a long-time average, therefore, speech is a low-frequency noise. Most of the power in the speech-wave is carried by frequencies below 1000 cycles. Thus, even in ignorance of the ear's susceptibility, the nature of speech itself suggests that communication is most seriously disrupted by sounds which have their energy concentrated in the frequencies below 1000 cycles.

ARTICULATION TESTING

When the over-all pressure of the speech-wave is less than about 5-10 db re 0.0002 dyne/cm², the speech is inaudible. As the intensity of the speech is gradually increased the presence of spoken sounds can be detected even though none of the words is distinguishable. This intensity is conveniently referred to as the *threshold of detectability* for speech (3). Approximately 8 db above the threshold of detectability, the sounds begin to be perceived as words. If the spoken material is continuous discourse, it is possible, though difficult, to understand the gist of the

* A recent attempt to include the temporal dimension is reported by R. K. Potter (15).

passage. This intensity is called the *threshold of perceptibility*. If the intensity is increased about 4 db more, the listener is able to obtain without perceptible effort the meaning of almost every sentence and phrase of the connected discourse. Above this level, the *threshold of intelligibility*, the intensity of the speech can be increased until it becomes painful at about 140 db (1). These thresholds are distinct and reliable, and different listeners will agree on their value.

Any one of the three thresholds can be used to determine the shift in the threshold due to the presence of a masking sound. With listeners instructed to determine the threshold of perceptibility in the presence of noises of variable intensity, functions of the type shown in Fig. 2 are obtained. Thus a pure tone of 1000 cycles at an intensity of 100 db shifts the threshold of perceptibility 18 db. A tone of 300 cycles at 100 db shifts the threshold 42 db, and a random noise of 100 db produces a 68-db shift.

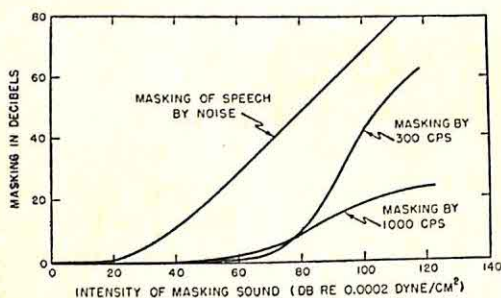


FIG. 2. THE SHIFT IN THE THRESHOLD OF PERCEPTIBILITY FOR SPEECH AS A FUNCTION OF THE INTENSITY OF DIFFERENT MASKING SOUNDS.

Random noise and pure tones having frequencies of 300 and 1000 cycles were used to mask the speech.

For a precise determination of the threshold, however, more elaborate methods are available. A talker reads out a list of discrete words, and listeners record what they hear. The percentage of the words heard correctly is taken as the articulation score. This method stands in the same relation to the threshold-methods as does the method of constant stimuli to the method of average error in classical psychophysics. By changing the intensity of either the speech or the masking sound a series of articulation scores ranging from 0% (no words heard) to 100% (all words heard correctly) is obtained. The 50% point on this function can be regarded as the threshold.

It is obvious that the results of these tests depend upon the type

of spoken material, the proficiency of talker and listeners, the characteristics of the equipment, etc. During the course of the development of articulation testing methods, from their first use by the Bell Telephone Laboratories (4) to their present wide application in communication problems, many variations have developed and considerable work has been devoted to their evaluation and standardization (3). The method adopted in the research reviewed here is recommended by its economy of time and personnel. Talkers are replaced by their recorded voices. Lists of difficult words are phonographically recorded in several different scramblings by each of five talkers. Between two and four trained listeners are used. As the records are played, the listeners follow the words with the aid of a check-list, uncovering each word *after* they hear it spoken. The listeners indicate by check-marks or manual counters whether or not they hear the word correctly. Since each listener must establish his own criterion for "hearing" a word, training is necessary before consistent results are obtained. A comparison with results from more formal articulation tests, however, indicates that with trained listeners the abbreviated procedure gives valid articulation scores.

It will be noted that the masking produced by various noises can be compared without obtaining the complete masking functions illustrated in Fig. 2. Single points on each function, if properly chosen, indicate the relative disruption of vocal communication which the different noises produce. Thus, if a quick check is desired, the threshold of perceptibility or intelligibility can be determined. If a more accurate value is needed, complete articulation functions can be run. One method is useful when a large number of variables need to be surveyed; the other gives more precise information when the significant variables have already been determined.

MASKING SPEECH BY TONES

Sounds which occur around us every day are customarily classified as either tones or noises. In this usage the word "tone" implies a harmonic relation between component frequencies. The word "noise" indicates that the component frequencies are dissonant, or randomly related. This is a classification of convenience, and a rigorous attempt to apply it to all the sounds which can be produced would quickly reveal borderline cases. Human speech, for example, is composed of sounds which have tonal characteristics (e.g. the vowels) and sounds with noisy characteristics (e.g. unvoiced plosives and fricatives), as well as many sounds which have both tonal and noisy components. In order to classify the variety of sounds which might interfere with

speech, at least three categories are required: (1) tones, (2) noises, and (3) voices. This rather arbitrary classification of the various sounds will serve to organize the following discussion.

Pure tones. As we would expect from the characteristics of speech, and as was indicated in Fig. 2, tones of low frequency produce more interference with speech than do tones of high frequency. This fact is supported by a very thorough study reported by Stevens, Miller and Truscott (21). These investigators found that for weak intensities the maximal masking is produced by sine waves in the vicinity of 500 cycles, whereas at high intensities the greatest masking occurs near 300 cycles. This effect is presumably due to the rapid spread of masking into the higher frequencies as the intensity of the masking tone is increased (22). This upward spread is aided by distortion in the ear which produces aural harmonics at high sound-levels (4, 20), and also by the characteristics of the response of the individual auditory-nerve fibers (9, 10, 13).

Complex Tones. Stevens, Miller and Truscott also determined the masking produced by square waves and by repetitive pulses of 10-microsecond duration. Square waves are less critical than sine waves as to frequency; fundamental frequencies between about 80 and 400 cycles mask with approximately equal effectiveness. No shift in the optimal frequency was observed as the intensity of the complex tones was increased. If the peak amplitude of the pulses is held constant, the greatest interference is produced by a pulse-repetition-frequency of about 200 pulses per second.

When the various wave-forms are equated for sound-pressure level, the sine wave is about 7 db less effective than the square wave, and the square wave about 7 db less effective than the pulses. This comparison is made for the low-frequency range in which the tones are most effective. At frequencies above about 1000 cycles all the tones are equally ineffective.

Before proceeding to discuss other types of interference we should pause to note two important generalizations supported by these data. First, low frequencies are more effective than high, and second, the greater the harmonic content of the tone the better its spectrum blankets the frequency-range of speech.

Patterns of Tones. Essentially the same considerations apply to the case of tones which are changing in frequency. Tests have been conducted with signals which changed abruptly or gradually in frequency, which covered wide or limited frequency-ranges at slow, intermediate or rapid rates, which varied in complexity or were accompanied by

steady tones, etc. In all the variations of tonal signals, the significant dimensions for masking were the frequency of the fundamental and the richness of the harmonic content.

Some examples should document these statements. One series of experiments was conducted with warbling tones—the tone rose slowly from the lowest to the highest frequency, then dropped suddenly back to the lowest frequency and rose again. The tone was produced by a relaxation oscillator and was rich in harmonics. Such a warble can be varied in center-frequency, range or rate. With the speech (heard binaurally

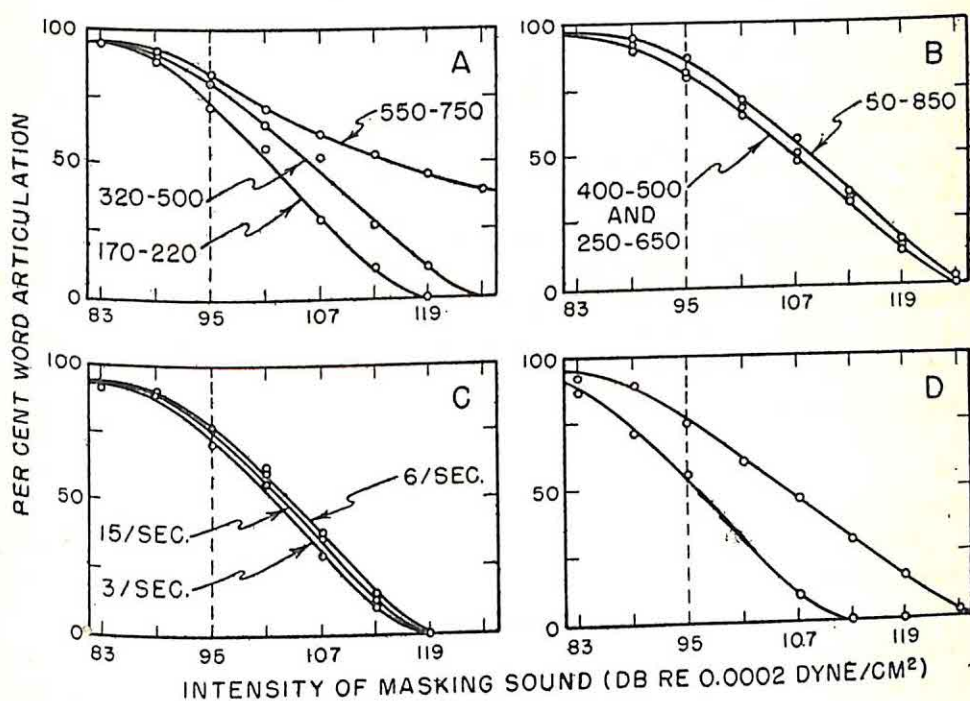


FIG. 3. THE PERCENT OF THE WORDS WHICH WERE CORRECTLY HEARD AS A FUNCTION OF THE INTENSITY OF VARIOUS TONAL MASKING SOUNDS.

Different center-frequencies of a warbling sound are compared in A; different ranges in B, different rates in C. In D the effect of adding a steady tone to a stepped pattern of tones is illustrated. The level of the speech was held constant at 95 db.

in dynamic earphones PDR-3) held at 95 db, the intensity of the warble was raised in 6-db steps until the functions shown in Fig. 3 had been obtained. Fig. 3-A compares warbles which had different center-frequencies. The high-pitched warble (550 to 750 cycles) resulted in the least interference and 50% of the words were still intelligible when

speech, at least three categories are required: (1) tones, (2) noises, and (3) voices. This rather arbitrary classification of the various sounds will serve to organize the following discussion.

Pure tones. As we would expect from the characteristics of speech, and as was indicated in Fig. 2, tones of low frequency produce more interference with speech than do tones of high frequency. This fact is supported by a very thorough study reported by Stevens, Miller and Truscott (21). These investigators found that for weak intensities the maximal masking is produced by sine waves in the vicinity of 500 cycles, whereas at high intensities the greatest masking occurs near 300 cycles. This effect is presumably due to the rapid spread of masking into the higher frequencies as the intensity of the masking tone is increased (22). This upward spread is aided by distortion in the ear which produces aural harmonics at high sound-levels (4, 20), and also by the characteristics of the response of the individual auditory-nerve fibers (9, 10, 13).

Complex Tones. Stevens, Miller and Truscott also determined the masking produced by square waves and by repetitive pulses of 10-microsecond duration. Square waves are less critical than sine waves as to frequency; fundamental frequencies between about 80 and 400 cycles mask with approximately equal effectiveness. No shift in the optimal frequency was observed as the intensity of the complex tones was increased. If the peak amplitude of the pulses is held constant, the greatest interference is produced by a pulse-repetition-frequency of about 200 pulses per second.

When the various wave-forms are equated for sound-pressure level, the sine wave is about 7 db less effective than the square wave, and the square wave about 7 db less effective than the pulses. This comparison is made for the low-frequency range in which the tones are most effective. At frequencies above about 1000 cycles all the tones are equally ineffective.

Before proceeding to discuss other types of interference we should pause to note two important generalizations supported by these data. First, low frequencies are more effective than high, and second, the greater the harmonic content of the tone the better its spectrum blankets the frequency-range of speech.

Patterns of Tones. Essentially the same considerations apply to the case of tones which are changing in frequency. Tests have been conducted with signals which changed abruptly or gradually in frequency, which covered wide or limited frequency-ranges at slow, intermediate or rapid rates, which varied in complexity or were accompanied by

steady tones, etc. In all the variations of tonal signals, the significant dimensions for masking were the frequency of the fundamental and the richness of the harmonic content.

Some examples should document these statements. One series of experiments was conducted with warbling tones—the tone rose slowly from the lowest to the highest frequency, then dropped suddenly back to the lowest frequency and rose again. The tone was produced by a relaxation oscillator and was rich in harmonics. Such a warble can be varied in center-frequency, range or rate. With the speech (heard binaurally

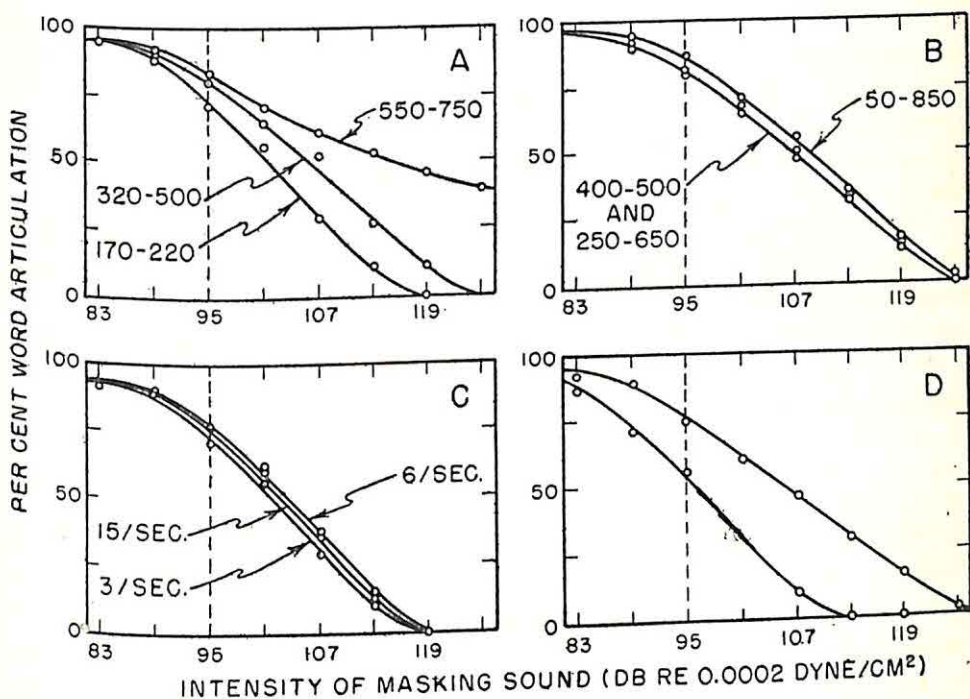


FIG. 3. THE PERCENT OF THE WORDS WHICH WERE CORRECTLY HEARD AS A FUNCTION OF THE INTENSITY OF VARIOUS TONAL MASKING SOUNDS.

Different center-frequencies of a warbling sound are compared in A; different ranges in B, different rates in C. In D the effect of adding a steady tone to a stepped pattern of tones is illustrated. The level of the speech was held constant at 95 db.

in dynamic earphones PDR-3) held at 95 db, the intensity of the warble was raised in 6-db steps until the functions shown in Fig. 3 had been obtained. Fig. 3-A compares warbles which had different center-frequencies. The high-pitched warble (550 to 750 cycles) resulted in the least interference and 50% of the words were still intelligible when

the warble was 20 db more intense than the speech. The low-pitched warble (170 to 220 cycles) interfered most, and 50% of the words were unintelligible when this sound was only 7 db more intense than the speech.

In Fig. 3-B three warbles are compared which had the same rate (1.5/sec.) and the same center-frequency (450 cycles), but which differed in the range of frequencies covered by the warbling tone. The widest range (50 to 850 cycles) was slightly less effective because it included some relatively high frequencies, but the differences are quite small.

Different rates of warble are compared in Fig. 3-C, and no significant change in masking results from varying the rate of modulation between 3 and 15 warbles per second.

Fig. 3-D presents results for a slightly different type of tonal interference. In this case the complex tones varied in frequency by abrupt steps, and an irregular pattern of 5 tones ranging in fundamental frequency between 300 and 600 cycles was repeated over and over. The masking produced by this signal is indicated by the right-hand function in Fig. 3-D. A steady, complex tone of 200 cycles was then added to the stepped tones, and the left-hand function resulted. The improved coverage of the component frequencies of speech by this more complex sound increased the masking by about 10 db (for 50% word articulation).

These results, which are typical samples drawn from a large body of similar data, show that the important aspects of these sounds are the aspects related to the adequate coverage of the speech-spectrum. Variations in the characteristics of the masking sound which do not affect this coverage do not affect the masking.

Music. Since much of the popular dance music of the day is (to some people) noisy and annoying, the possibility that it interferes seriously with speech was worth investigating. Listening tests quickly revealed, however, that most music is inoffensive for two reasons. Music is divided into phrases, and between phrases are pauses, and during pauses speech is intelligible. And in the second place, clarinet and trumpet solos usually fall above the range of frequencies which produces the most efficient masking of speech.

A very complex masking sound is obtained, however, if two or three phonographic recordings are played at the same time. The different orchestras fill in each other's pauses, and the coverage of speech-frequencies is more consistently adequate. With a signal of this type, an articulation function was obtained which was almost identical with the

function in Fig. 3-D for the complex stepped tones. Presumably, this function represents nearly the maximum masking which a tonal signal can produce. If the sound is made more complex it loses its tonal characteristics and begins to sound more and more like an irregular noise.

MASKING BY NOISE

In many ways white noise, having a uniform, continuous spectrum, is the ideal sound to use in studying the masking of speech. This noise provides a continuous coverage of a wide range of frequencies, and the spectrum can be manipulated by filters to suit the manipulator's fancy.

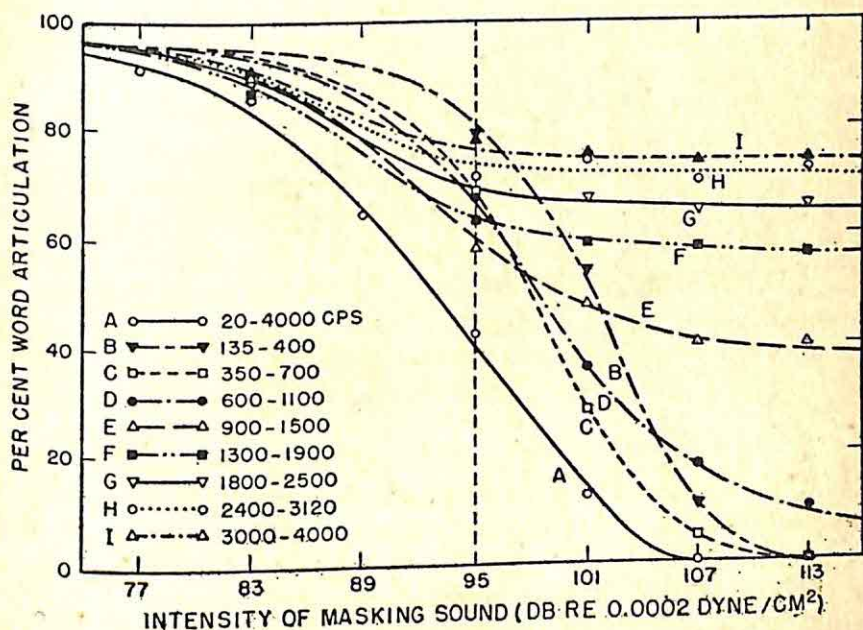


FIG. 4. THE ARTICULATION SCORE AS A FUNCTION OF THE INTENSITY OF THE MASKING NOISE.

The different frequency-bands of noise are parameters. The level of the speech was held constant at 95 db.

Thus we might use wide or narrow bands of noise, or noises "shaped" for maximal effect.

Narrow Bands of Noise. Let us review first some results obtained with narrow bands of noise. Highly discriminating filters (M-derived) were inserted into the noise-channel to provide adjacent pass-bands as follows: 135-400 cycles, 350-700 cycles, 600-1100 cycles, 900-1500 cycles, 1300-1900 cycles, 1800-2500 cycles, 2400-3120 cycles, and 3000-4000 cycles. These cut-off frequencies represent the frequency at which

the response of the system was 6 db below the maximum in the band. The narrow bands of noise were then mixed with the speech, which was held constant at 95 db, and articulation tests were run. The results are shown in Figs. 4 and 5. In Fig. 4 the articulation score is plotted as a function of the noise-level, with the different bands as parameters. In Fig. 5 the articulation score is plotted as a function of the center-frequency of the bands of noise (vertical divisions indicate the approximate cut-off frequencies of the bands), with the noise-level as the parameter.

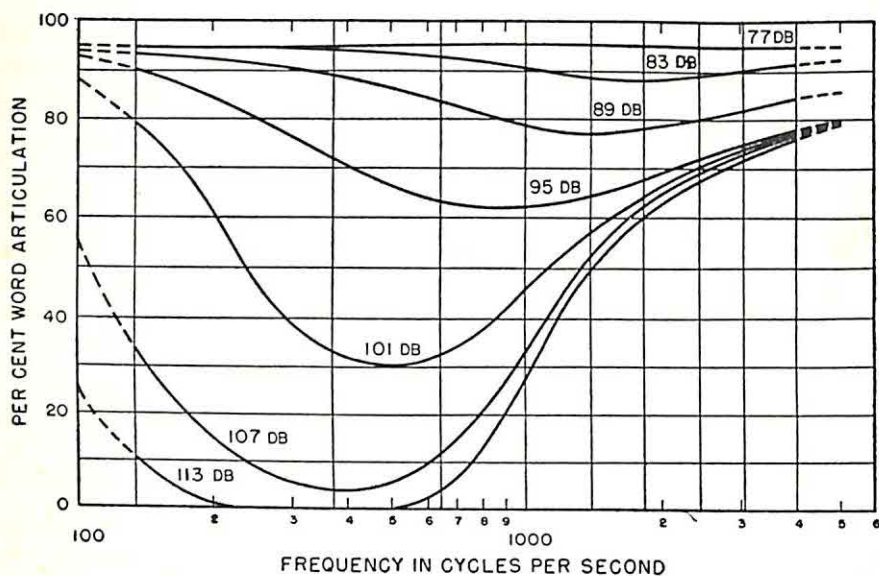


FIG. 5. SMOOTHED CURVES DRAWN FROM THE DATA OF FIG. 4, SHOWING THE ARTICULATION SCORE AS A FUNCTION OF THE COMPONENT FREQUENCIES OF THE MASKING NOISE FOR DIFFERENT NOISE-LEVELS.

Complete masking of the speech was obtained only with the two lowest bands. With bands of noise above 1000 cycles the speech could be heard even when the noise was more than 18 db above the level of the speech. Results with a wide band of noise (20 to 4000 cycles) are also shown in Fig. 4, and it will be noted that none of the narrow bands of noise produced more masking than the unfiltered noise. At the low noise-levels, the high-frequency bands were more effective than bands below 1000 cycles. At the high noise-levels, however, the low-frequency bands were more effective. This is seen most clearly in Fig. 5.

In order to interpret these results it is necessary to recall the results of studies on the masking of pure tones by pure tones. Sounds of low frequency will, if intense enough, eventually mask the entire range of

frequencies involved in speech, but the high-frequency sounds are not able to mask the low frequencies of speech. The fact that, at low noise-levels, the high-frequency bands of noise are more effective than the low is probably due to the distribution of energy in speech. The higher frequencies of speech (cf. Fig. 1) are much weaker and are therefore more easily masked. As the intensity of the high-frequency noise is raised, however, the high-frequency sounds of speech (mostly unvoiced consonants) are masked, but this masking does not spread to the low-frequency sounds. If, on the other hand, a band of noise composed of low frequencies is gradually increased in intensity, it will at first have little effect because it is not so strong as the low speech-frequencies, nor is it intense enough to spread its masking effects to the higher, weaker speech-sounds. As the intensity is further increased, however, the masking effects begin to cover the entire frequency-range of speech, and complete masking is rapidly obtained.

It follows from these results that the most masking is produced by a spectrum with the noise-energy in the lower third of the range from 100 to 4000 cycles. Experimentation with a wide variety of "tailored," sloping noise-spectra showed that the optimal masking noise has a spectrum similar to the long-interval speech-spectrum which is being masked.

Modulation of the Noise. Random noise is often described as a steady "hissing" sound which, at low levels, is not unpleasant to hear. Unlike intermittent bursts of static or the irregular clacking of a typewriter, the sustained nature of random noise enables the listener to adapt to it and, to a certain extent, to ignore it. Consequently, different methods were tried for changing or modulating the quality of the noise. One method used filters which were switched in and out irregularly. Another made use of the Sonovox, a device which can be substituted for the vocal cords as a sound-source for speech. With a Sonovox driven by a random noise voltage, the resulting "speech" resembles a hoarse breathy whisper.

The results with such noises need not be reviewed in detail. It is sufficient to say that the experimenters were again forced to the conclusion that the critical dimension of the masking noise is its spectrum. Modulations of the noise may affect the listener's comfort, for he will say he prefers to listen to the steady noise. But when he is driven to it, he will understand just as many words with a modulated masking noise as with a steady noise.

The Masking Wave-Form. Will two noises having identical spectra, but different wave-forms, produce the same masking effects?

In order to study this question a frequency-modulated (FM) audio oscillator was developed. This oscillator was modulated with noise, i.e., the frequency of oscillation was randomly varied. The resulting wave-form, shown schematically in the upper half of Fig. 6, resembles a sine wave of constant amplitude, but of varying frequency. Once the spectrum of the frequency-modulated noise had been measured, it was

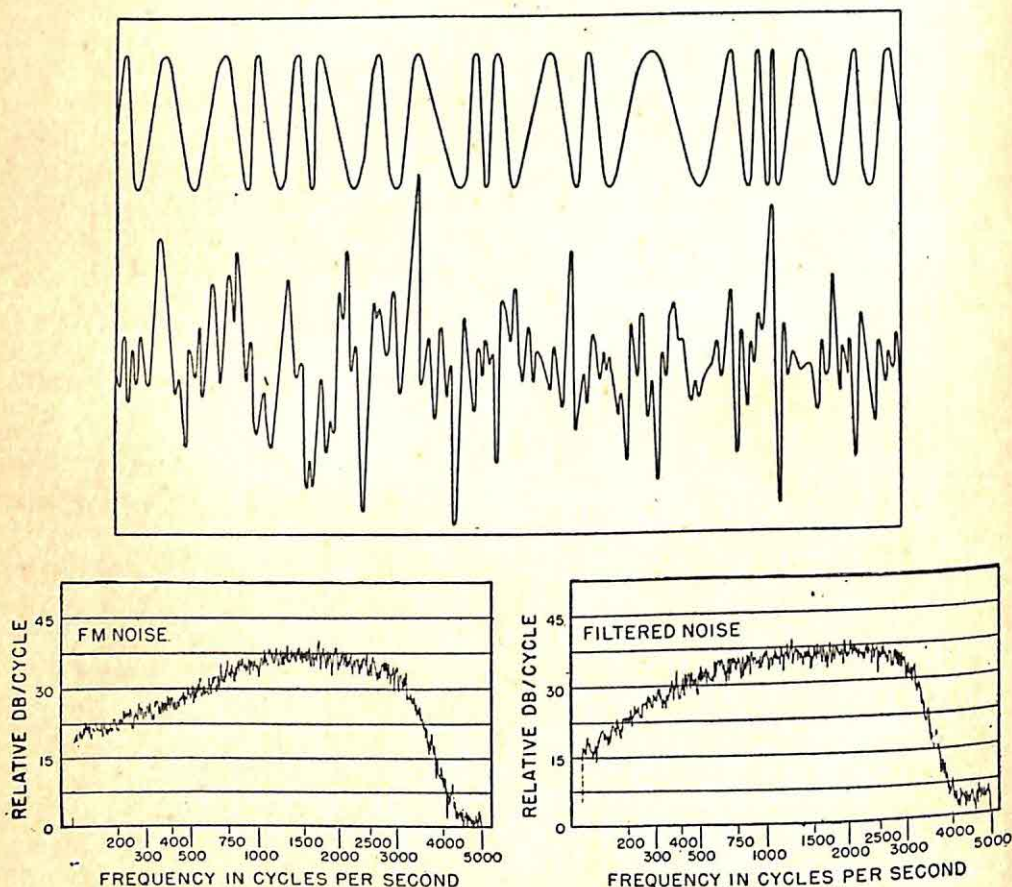


FIG. 6. FREQUENCY-MODULATED (FM) NOISE VS. RANDOM NOISE.

The wave-form at the top of the figure schematizes FM noise; the wave-form below, random noise. The matched spectra of FM and random noise are shown in the lower half of the figure.

then possible by the use of filters to "shape" the spectrum of a random noise of irregular amplitude in such a way as to make it correspond with the spectrum of the FM noise. When the two spectra shown in the lower half of Fig. 6 are compared, it is seen that the spectrum of the

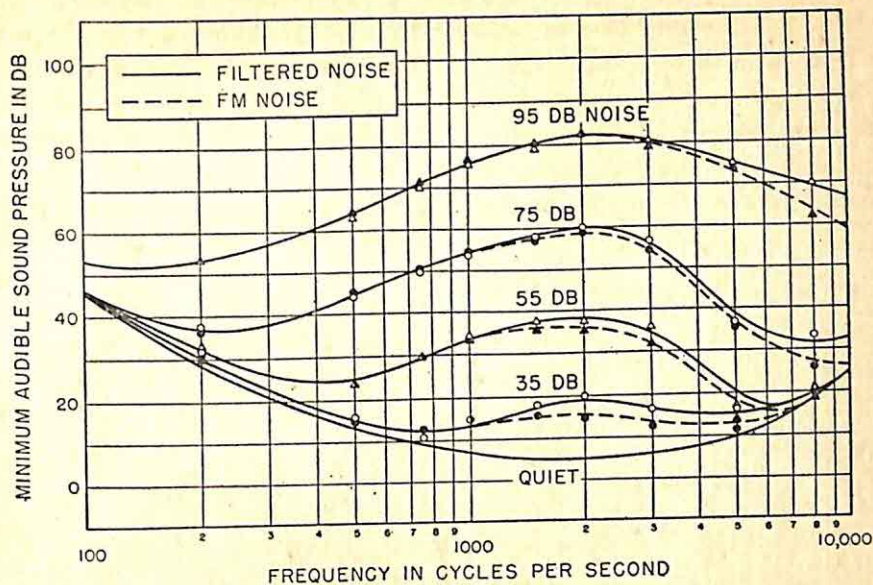


FIG. 7. THE SHIFT IN THE THRESHOLD OF AUDIBILITY PRODUCED BY TWO NOISES HAVING DIFFERENT WAVE-FORM BUT SIMILAR SPECTRA.

(Cf. Fig. 6.)

FM noise is 2 or 3 db lower at frequencies above 2000 cycles. Otherwise the spectra are very similar.

The masking by these two wave-forms was then tested for both pure tones and speech, and the results are shown in Figs. 7 and 8. In

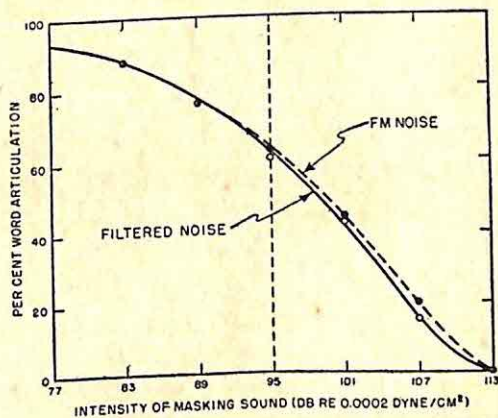


FIG. 8. THE ARTICULATION SCORE AS A FUNCTION OF THE INTENSITY OF TWO MASKING NOISES HAVING DIFFERENT WAVE-FORMS BUT SIMILAR SPECTRA.

(Cf. Fig. 6.) The level of the speech was held constant at 95 db.

Fig. 7 the masked thresholds are plotted against the frequencies of the masked tones, with the intensity of the noise as the parameter. The small discrepancy between the masking effectiveness of the two noises correlates with the fact that the FM noise contained slightly less energy at the higher frequencies. In Fig. 8 the articulation score is plotted against noise level with wave-form as the parameter. The two functions are not significantly different.

In the light of these results it is apparent that, within certain broad limits, the masking produced by a sound depends upon the spectrum of the sound and is independent of the phase-relations among the component frequencies. This conclusion was verified in another set of tests which compared random noise with the same noise after it had passed through a circuit (a "peak clipper") which sharply limited the instantaneous peak amplitudes of the noise voltage. Peak-clipped noise resembles a square wave randomly modulated in frequency, but its spectrum and masking effectiveness are the same as that of unclipped noise, provided the intensities are the same. Within broad limits, therefore, the actual form of the masking wave as seen on a cathode-ray oscilloscope is of no consequence in determining masking. This agrees with the general hypothesis that the ear is relatively insensitive to phase-relations. The possible exceptions arise when the masking spectrum is produced by very intense but very intermittent bursts of noise. With such a sound the recovery-time of the ear becomes of appreciable importance, and we will have to consider specifically these temporal factors a few pages hence.

MASKING BY OTHER VOICES

It has been said that the best place to hide a leaf is in the forest, and presumably the best place to hide a voice is among other voices. Certainly it is a practical problem, for the babble of other voices is a frequent background for our spoken words, and no survey of common masking sounds should overlook the noises which we ourselves produce. In considering "other" voices as sounds masking "the" voice, two considerations are obviously crucial. How many other voices is the speaker competing with? And what are the other people saying?

It is relatively easy for a listener to distinguish between two voices, but as the number of rival voices is increased the desired speech is lost in the general jabber. This fact is demonstrated in the articulation functions of Fig. 9. The groups of interfering talkers were composed of an equal number of men and women (the single masking voice was a man's) reading and talking in a conversational tone. The "desired"

voices were, in all cases, male voices. For convenience in testing, the babble of voices was phonographically recorded by a high-fidelity recording system.

Note that the single voice is a relatively poor masking signal, and that even two voices are less effective than four or more. Although the long-interval spectrum of a single voice is nearly optimal for masking speech, the spectrum at any moment does not include all of the neces-

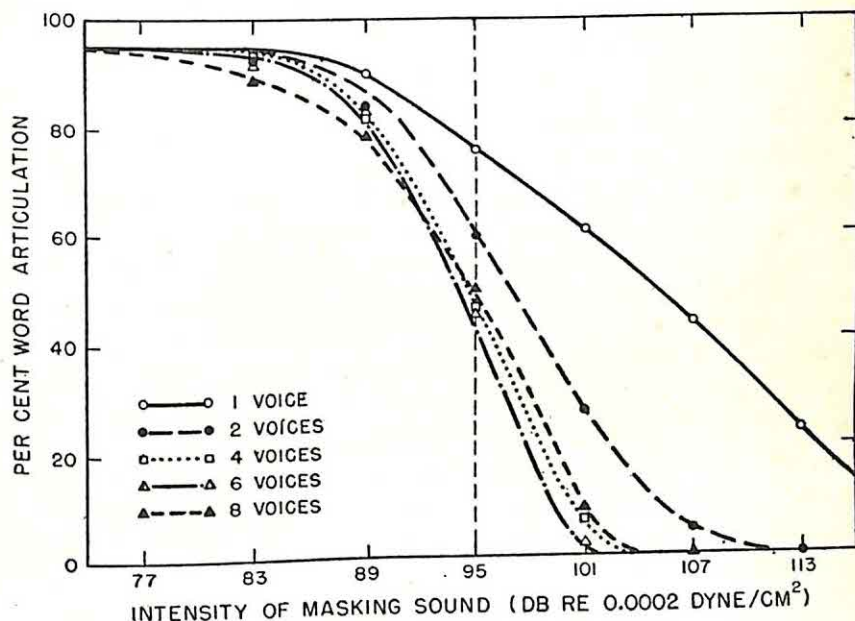


FIG. 9. THE ARTICULATION SCORE AS A FUNCTION OF THE INTENSITY OF DIFFERENT NUMBERS OF MASKING VOICES.

The level of the desired speech was held constant at 95 db.

sary frequencies. The variations in the level of a single voice are great, and there are relatively long intervals during which no masking sound is present. With several voices, however, a continuous masking signal is produced.

The content of the masking speech is a more difficult factor to evaluate. Conversational voices were compared with loud, excited voices liberally interspersed with laughter, cheering and improbable vocal effects. The two sounds could be likened to the chatter at a friendly dinner-party versus the din of a particularly riotous New Year's Eve celebration. There was little difference in masking, however. The shouting voices were a little more effective at low noise-levels, but

this was correlated with more high-frequency energy in the spectrum of the shouted babble.

Conversational babblings in different languages were also compared. A language was chosen which the listeners did not know, but the masking was neither greater nor less than was obtained with an English babble. Once again, it is necessary to conclude that the crucial factor is the masking spectrum. The particular way in which the spectrum is produced is of secondary importance.

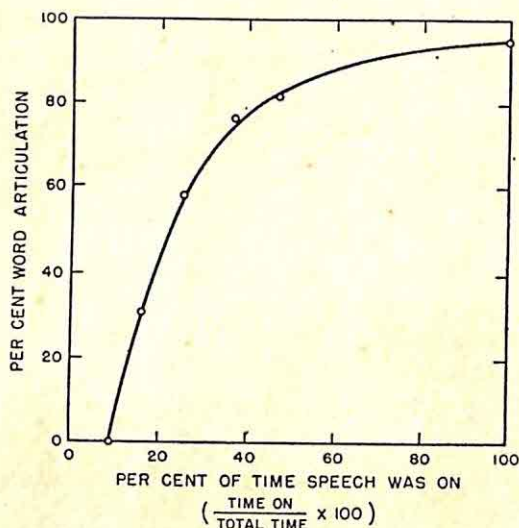


FIG. 10. THE ARTICULATION SCORE AS A FUNCTION OF THE PERCENT OF THE TIME THE SPEECH WAS ON.

The level of the speech was held constant at 95 db, and the speech-wave was interrupted at a rate of nine times per second.

TEMPORAL CONTINUITY OF MASKING SOUND

Not all of the sounds which force us to raise our voices are continuous⁴. They come and go, change in quality and loudness. Consequently, some attention should be paid to this dimension of the masking sound. Because the ear is facile in patching together interrupted fragments of speech, an intermittent noise is not as serious a hazard as a continuous noise.

In order to demonstrate that large portions of the speech can be completely blanked out without seriously lowering intelligibility, an electronic switch was used to interrupt the speech nine times a second for variable portions of the total time. The function in Fig. 10 shows the relation between the intelligibility of words and the percentage of

the time the speech was present. With the speech on only half the time, 80% of the words were still correctly heard. It was not until 90% of the speech wave was missing that none of the words could be understood. (It should be noted, of course, that the function of Fig. 10 depends upon the rate of interruption and the type of test material, as well as upon the on-off ratio.) Qualitatively, the interrupted speech seems "hoarse" or "husky," as if the talker had some disorder of phonation. Nonetheless, the speech is surprisingly intelligible.

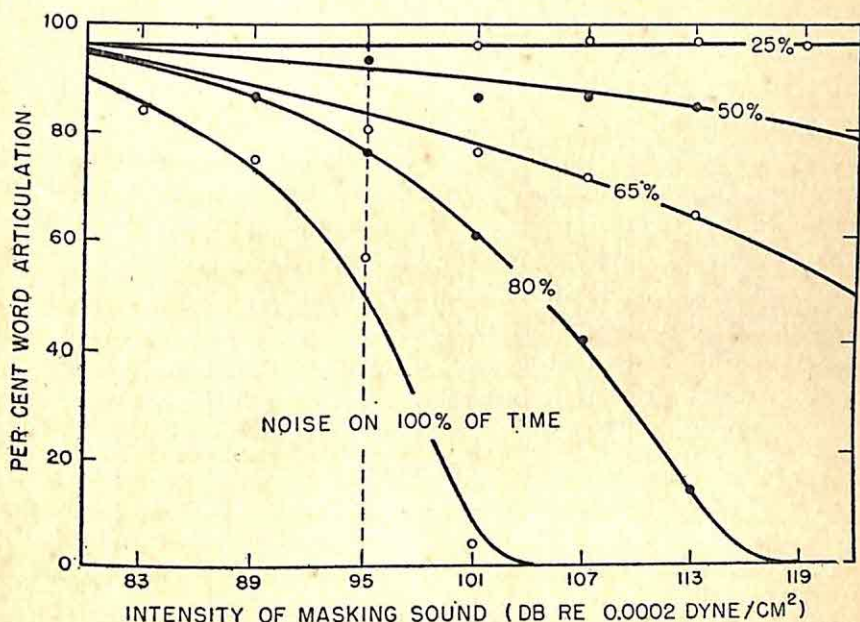


FIG. 11. THE ARTICULATION SCORE AS A FUNCTION OF THE INTENSITY OF INTERRUPTED MASKING NOISES.

The percent of the time the noise was on as parameter. The level of the speech was held constant at 95 db.

These results suggest that with an interrupted masking sound the ear should be able to integrate and interpret those portions of the speech available when the masking sound is not present. Consequently, the electronic switch was used to interrupt a masking noise instead of the speech. The results are shown in Fig. 11, where the articulation score is plotted against the intensity of the masking noise, with the percentage of the time the noise was on as the parameter. The noise-level in this case is expressed in terms of the intensity of the noise when it is on, and no average is taken between the level of the noise when it is on and the level when it is off.

When the noise is present less than 50% of the time, practically no masking is produced. Noise present 80% of the time is considerably less effective than uninterrupted noise. Apparently the recovery of the ear is rapid enough, and our ability to integrate fragments of speech is great enough, that any periodic interruption of a masking sound lowers its masking effectiveness. The careful documentation of this conclusion, however, remains for future investigators.

SPEECH-TO-NOISE RATIO

Articulation testing is a tedious and expensive business. The direct determination of the masking produced by a particular noise in some practical situation is not always feasible. Consequently considerable attention has been devoted, especially at the Bell Telephone Laboratories (8, 14), to the development of a computational procedure for predicting articulation scores. The simplest of the various proposed procedures relate the intelligibility of the speech sounds to the speech-to-noise ratios in the different frequency-ranges of speech.

The speech-to-noise ratio, as the words imply, is the ratio in decibels of the speech-intensity to the noise-intensity. It is usually convenient to distinguish two different but related ratios: (1) the over-all speech-to-noise ratio, and (2) the speech-to-noise ratio per cycle. To determine the over-all ratio the rms levels of speech and noise are measured directly and expressed as a ratio. Thus, for example, with speech at 95 db and a masking sound at 90 db, the over-all speech-to-noise ratio is +5 db.

We have seen, however, that masking depends upon the distribution of the component energy over the frequency-range. Two noises with the same over-all level can produce quite different masking results depending upon the spectra of the noises. To estimate masking effectiveness, therefore, it is necessary to know the spectrum of the masking noise and the spectrum of the speech. With an audio-spectrometer the two spectra are determined. The ratio in decibels of the speech-level-per-cycle to the noise-level-per-cycle at any frequency is the speech-to-noise ratio for that frequency. Thus the speech-to-noise ratio per cycle varies with frequency, while the over-all ratio does not consider the spectral distribution of the two sounds.

Now in order to use speech-to-noise ratios in the computation of articulation scores, it is necessary to make three assumptions. These can be considered in order.

1. *Articulation scores depend upon the type of test-material used.* With any fixed speech-to-noise ratio, higher scores will be obtained with sentences than

with words, and the lowest scores will be obtained with meaningless syllables. It is desirable, therefore, to obtain some more fundamental measure independent of the particular test-material. Sentence, word, or syllable scores could then be related to this more fundamental index, and all three could be predicted directly once the fundamental measure had been computed. Such a fundamental measure has been proposed by the Bell Telephone Laboratories. They call it the *articulation index*, and its derivation leads us to the second assumption.

2. *The contribution to the articulation index made by any narrow band of speech frequencies is independent of the contribution made by other bands.* If the range of speech-frequencies could be divided into a number of narrow bands each contributing the same amount to intelligibility, and each contribution independent of all other contributions, the sum of the individual contributions could be taken as the articulation index.

The Bell Telephone Laboratories attempted to establish empirically the relation between frequency and intelligibility by articulation tests conducted with filtered speech. With male voices, for example, they found that frequencies of speech above 1660 cycles gave the same articulation score (68% with the test material they were using) as the frequencies below 1660 cycles. For that crew and those test-materials, therefore, an articulation score of 68% was equivalent to an articulation index of 0.50, and 1660 cycles divided the speech into two bands, one above and one below, which contributed equally, 0.50 apiece, to the total index. In a similar manner the range of frequencies was further subdivided until a complete function could be drawn relating the articulation index to frequency. Once this function was determined, the total range of speech frequencies was divided into a convenient number of bands, each contributing equally to the articulation index (19). When the intensity of the speech in all the bands is far enough above threshold, the total contribution of all bands gives an articulation index of 1.00. Just how far above threshold the speech in any band should be brings us to the third assumption.

3. *The contribution to the articulation index made by any narrow band of speech-frequencies depends upon the speech-to-noise-ratio in that band.* If the speech in a given band of frequencies falls far below the level of the noise in that band, the speech will make no contribution to the articulation index. It is necessary, therefore, to use a weighting value to express the fractional part of the maximum contribution the band can make. This weight depends directly upon the speech-to-noise ratio. When the speech is 30 db or more above the noise, the band makes its maximum contribution, and speech-to-noise values greater than 30 db are considered as equivalent to the maximum weighting.

The computational procedure evolving from this argument is straightforward. First we determine the spectra of the speech and of the noise. Then we divide the range of speech-frequencies into n bands. The weighting value for each band of frequencies is next computed from the speech-to-noise ratios in the bands. The maximum contribution of every band is multiplied by its weighting factor. And then the contributions of all the bands are totalled to give the articulation index. Finally, the index is converted into sentence, word, or syllable articulation scores as desired.

Obviously this description of the procedure is highly simplified, and many special cases arise. For example, when narrow bands of noise are used to mask speech, the procedure does not provide adequate estimates of the masking effectiveness of the low-frequency bands of noise. Suppose that we have divided the speech spectrum into 20 equivalent bands. According to the computational procedure outlined above, the narrow band of noise will mask out 4 or 5 of the 20 bands, but the remaining 15 or 16 bands will continue to contribute to the articulation index regardless of the noise-level in the masked bands. Articulation scores should, according to prediction, fall a few percentage points and then remain constant for higher and higher noise-levels. Actually, as Fig. 4 shows, this is the case for high-frequency bands of noise. For low-frequency bands, however, there is a shift of masking into the frequencies above the band, and this spread increases as the intensity is increased. A similar example could, of course, be drawn from the results obtained when pure tones are used to mask speech.

This particular difficulty can, perhaps, be encompassed in a computational procedure if the masked threshold is used instead of the noise-spectrum. But in order to restrict the process entirely to pencil and paper, it is necessary to be able to predict the masked threshold for tones from the spectrum of the noise. This prediction can be made satisfactorily for noises with continuous spectra (5, 6, 7), but the prediction for tones or narrow bands of noise cannot be made at present.

Another problem which was encountered in evaluating masking noises was the matter of temporal continuity. The procedure outlined does not consider interrupted noises, even though interruptions are crucial to masking effectiveness. If more were known of the build-up and recovery times of the ear, this dimension of masking could presumably be quantified in a satisfactory way. For the present, however, the temporal aspect of the masking process has been neglected.

In spite of these difficulties, a simple computational procedure has much to offer. As a first approximation its usefulness cannot be denied. The emphasis is placed upon the distribution of masking energy relative to the distribution of the energy in speech and a survey of masking noises confirms the belief that these spectral distributions are of central significance. Problems of wave-form, of patterning and modulation, of familiarity with the masking sound are ignored, and the results presented here indicate that these aspects of the masking sound are of negligible consequence.

The results of the research and of the discussion can therefore be briefly summarized: the greatest interference with vocal communication

is produced by an uninterrupted noise which provides a relatively constant speech-to-noise ratio over the entire range of frequencies involved in human speech. Unfortunately, most of the noises we compete with fill this general prescription.

THE QUESTION OF ANNOYANCE

It is all well and good to know about the masking effectiveness of unwanted sounds, but few of us spend much time in boiler-rooms, airplane-cockpits, or machine-shops, and when we do ride the subway we simply shut up until the car stops. Noises which we cannot shout out may be a hazard in some occupations, but most of us manage to get along fairly well in the average din of the average day. What really upsets our vocal communication is the sound that, although it does not shatter our eardrums, is a nuisance, that distracts or annoys us, or just "gets on our nerves." Every teacher knows the shudder that runs through a class when the chalk squeaks piercingly against the blackboard, and an otherwise amusing radio program can become an agonizing nuisance if we try to use the telephone.

When we take this problem into the laboratory, however, it seems to disappear right in front of our ears. The major difficulty rests in the fact that the listener's attitude is so important. If he is engaged in difficult mental work, it may be relatively easy to annoy him. But if he listens with a defiant attitude, any attempts to upset him with strange noises may prove more amusing than effective. And since most of the sounds we can use in the laboratory are out of context and relatively meaningless, the task of being successfully obnoxious is practically impossible. Annoyance depends primarily upon the particular listener and the particular situation in which he finds himself.

If, however, we are content to ignore some of the situational variables involved, it is possible to ask listeners to compare different sounds on the basis of their "annoyance value." Some simple listening situation is standardized and the listeners compare pairs, use a rating scale, or rank-order an array of sounds. One can then evaluate the variables contributing to annoyance value as defined by the situation, although the safety with which the results can be extended to other situations is open to question.

Listeners were presented with pairs of sounds and were instructed to indicate which of the two sounds was more annoying. In making this decision, the listeners were told to judge which sound of the pair would be more unendurable if they had to listen to it for a long period of time. These instructions, therefore, constitute the definition of annoyance.

Fairly consistent results were obtained with groups of 10 to 20 listeners, and the scale of annoyance constructed in this way agreed closely with results obtained with rating or rank-ordering procedures.

As an illustration, the results obtained with stepped patterns of tones will be considered. Eight different variables in the tonal pattern were studied for their effect upon annoyance-value.

1. The higher the pitch of the component tones, the greater the annoyance-value. The range of frequencies tested was from 200 to 1500 cycles.

2. A wide range of frequencies between the highest and lowest steps is more annoying than a restricted range. Listeners reported that the wide range of component frequencies tended to be perceived alternately, first as a complete pattern and then as two patterns, one of high and one of low pitch. This effect is very similar to figure-ground reversals in visual perception.

3. The addition of continuous tones to the stepped pattern of tones produces complex effects dependent upon the frequency-relation between the tones. Beats give the sound a rough pulsing irregularity which the listeners disliked.

4. Listeners asked to compare continuous sounds of different wave-shapes found the complex sounds, especially brief pulses, more annoying. In general, the sine wave was found to produce little annoyance.

5. Patterns of 3, 4, 6 and 12 tones were compared, but the number of different steps in the complete pattern had little effect on the judgments of annoyance.

6. If one of the steps of a pattern is slightly longer in duration than the others, a rhythmic quality is added which the listeners judged to be more annoying than tones of equal duration. Even more annoying, however, is the pattern in which all the tonal durations are randomly varying.

7. A slow rate of repetition for a pattern of tones is considered slightly more annoying than a rapid rate.

8. Up to a certain limit, the annoyance-value is increased if silent intervals are introduced between the successive steps.

These results typify listeners' responses to meaningless sounds. When meaningful sounds like speech or music were used, the listeners refused to apply the word "annoyance" in describing them. Annoyance did not seem to be a proper dimension of such sounds, but the listeners were agreeable to calling the sounds "distracting." Apparently, meaningful sounds have a higher "attention-value" than meaningless sounds.

These experimental results, supplemented by results obtained with other types of sound, indicate that annoyance is related to three aspects of the sound.

Loudness. The most important single factor in determining annoyance-judgments is the intensity of the sound. With sufficient intensity, any sound can be made annoying, and extremely loud sounds produce actual pain. Since this variable is so fundamental to annoyance, care was taken to equate the intensity of the signals when other aspects were being studied.

Pitch. In general, sounds having their energy concentrated among the higher

audible frequencies are more annoying than low-frequency noises (11, 16). In this respect, the frequency of the sound alters its annoyance-value in a manner opposed to the effect on masking. With a low-frequency noise we cannot hear speech, but with a high-frequency noise we are more apt to be annoyed.

Modulation of Loudness and Pitch. A third important factor is the modulation which the sound undergoes. Listeners report that they prefer to listen to continuous, unchanging sounds, and that a sound changing irregularly from moment to moment is more annoying than a sound which is changing regularly (12). Listeners feel that the distraction of a changing sound is less desirable than the boredom of a constant sound, and they retain this opinion even after many hours of articulation testing in the presence of different noises. Apparently the changes in loudness are more effective than changes in pitch, but the individual differences on this point are too conspicuous to permit a safe generalization.

These conclusions lead to the belief that what we are here considering as annoyance is a close relative to the problem of hedonic tone. Judgments of the pleasantness, indifference, and unpleasantness would probably have led to very similar conclusions. Thus, while the results may be interesting as an exercise in experimental esthetics, the character of the problem has somehow been altered by the experimental approach. The principal concern, it will be recalled, is with annoyance as a hazard to vocal communication. On this score the results are consistently negative, and at no point in the experimental results is there unequivocal evidence that the articulation scores obtained by trained listeners in the presence of an annoying sound were lower than the scores obtained in the presence of an indifferent sound which had the same acoustic spectrum. With the attitude adopted by listeners in the laboratory situation, annoyance is not a hazard to communication. And yet, sounds do differ in annoyance value, and annoyance or distraction does sometimes interrupt our verbal flow. Perhaps the most reasonable generalization, therefore, is that when a listener finds himself in a situation where he is vulnerable to auditory annoyance, he is most vulnerable to loud, high-pitched, unpredictable sounds. Just what situational and attitudinal factors contribute to his vulnerability, however, this research does not reveal.

SUMMARY

A wide variety of sounds have been investigated to determine the extent to which they interfere with vocal communication. The masking of speech has been determined by articulation testing methods, and estimates of annoyance have been obtained by the method of paired comparisons.

The sounds are classified as noises, tones, and voices. For all three



types of sound, the stimulus-dimensions determining both masking and annoyance are the intensity, the frequency or spectrum, and the temporal pattern of the sound. Masking depends primarily on the speech-to-noise ratio over the range of frequencies involved in speech. Sounds of low frequency mask this range more effectively than sounds of high frequency. Interruptions in the sound decrease the masking effectiveness.

Annoyance also increases as the intensity is raised, but low-frequency sounds are less annoying than high-frequency sounds, and intermittent, irregular sounds are more annoying than continuous sounds. There is no evidence, however, that annoyance interferes with vocal communications in the laboratory situation.

BIBLIOGRAPHY

1. DAVIS, H., HUDGINS, C. V., MARQUIS, R. J., NICHOLS, R. H., PETERSON, G. E., ROSS, D. A., & STEVENS, S. S. The selection of hearing aids. *Laryngoscope*, 1946, 56, 85-115 (Report PNR-7 from the Psycho-Acoustic Laboratory, Harvard University).
2. DUNN, H. K., & WHITE, S. D. Statistical measurements on conversational speech. *J. Acoust. Soc. Amer.*, 1940, 11, 278-288.
3. EGAN, J. P. *Articulation testing methods II*. Psycho-Acoustic Laboratory, Harvard University, Nov. 1, 1944, OSRD Report No. 3802. (Available through the Office of Technical Services, U. S. Department of Commerce, Washington, D. C., PB 22848.)
4. FLETCHER, H. *Speech and hearing*. New York: D. van Nostrand, 1929.
5. FLETCHER, H., & MUNSON, W. A. Relation between loudness and masking. *J. Acoust. Soc. Amer.*, 1937, 9, 1-10.
6. FLETCHER, H. Loudness, masking and their relation to the hearing process and the problem of noise measurement. *J. Acoust. Soc. Amer.*, 1938, 9, 275-293.
7. FLETCHER, H. The mechanism of hearing as revealed through experiment on the masking effect of thermal noise. *Proc. nat. Acad. Sci., Wash.*, 1938, 24, 265-276.
8. FRENCH, N. R., & STEINBERG, J. C. Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Amer.*, 1945, 17, 103.
9. GALAMBOS, R., & DAVIS, H. The response of single auditory-nerve fibers to acoustic stimulation. *J. Neurophysiol.*, 1943, 6, 39-58.
10. GALAMBOS, R., & DAVIS, H. Inhibition of activity in single auditory nerve fibers. *J. Neurophysiol.*, 1944, 7, 287-304.
11. LAIRD, D. A., & COYE, K. Psychological measurements of annoyance as related to pitch and loudness. *J. Acoust. Soc. Amer.*, 1930, 1, 158-163.
12. LAIRD, D. A. Influence of noise on production and fatigue, as related to pitch, sensation level and steadiness of the noise. *J. appl. Psychol.*, 1933, 17, 320-330.
13. LOWY, K. Some experimental evidence for peripheral auditory masking. *J. Acoust. Soc. Amer.*, 1945, 16, 197-202.
14. MUNSON, W. A. Relation between the theory of hearing and the interpretation of speech sounds. *J. Acoust. Soc. Amer.*, 1945, 17, 103.
15. POTTER, R. K. Visible patterns of

- sound. *Science*, 1945, 102, 463-470.
16. REESE, T. W., & KRYTER, K. D. *The relative annoyance produced by various bands of noise*. Psycho-Acoustic Laboratory, Harvard University, March 17, 1944, IC-65.
17. RUDMOSE, H. W., et al. *Effects of high altitude on the human voice*. Electro-Acoustic Laboratory, Harvard University, Jan. 30, 1944, OSRD Report No. 3106. (Available through the Office of Technical Services, U. S. Department of Commerce, Washington, D. C., PB 19820.)
18. SIVIAN, L. J., & WHITE, S. D. On minimum audible sound fields. *J. Acoust. Soc. Amer.*, 1933, 4, 288-321.
19. STEINBERG, J. C., & FRENCH, N. R. The portrayal of visible speech. *J. Acoust. Soc. Amer.*, 1946, 18, 4-18.
20. STEVENS, S. S., & DAVIS, H. *Hearing*. New York: John Wiley, 1938.
21. STEVENS, S. S., MILLER, J., & TRUSCOTT, I. The masking of speech by sine waves, square waves, and regular and irregular pulses. *J. Acoust. Soc. Amer.*, in press. (Report PNR-14 from the Psycho-Acoustic Laboratory, Harvard University.)
22. WEGEL, R. L., & LANE, C. E. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Phys. Rev.*, 1924, 23, 266-285.
23. *American standard acoustical terminology*. American Standards Association. New York: Amer. Soc. Mech. Engineers, March 20, 1942.

THE EFFECTIVE USE OF MANIPULATIVE TESTS IN INDUSTRY

W. F. LONG AND C. H. LAWSHE, JR.

Division of Education and Applied Psychology, Purdue University

"The practical possibilities of psychological tests are now generally conceded, both by the professional psychologist and the industrial layman. . . . The immediate future is likely to see a very extensive application of tests to industry. . . ." These statements were made by Henry C. Link (35) in 1919.

It is disturbing to realize that those words seem just as apropos today as they did twenty-seven years ago, and that there is little doubt that the use of tests in industry during the intervening period until now has not been as extensive as Link anticipated. What are the reasons for this slow development of industrial testing, and what are the possibilities of Link's "immediate future" with the extensive application of tests to industry being close at hand? A review of pertinent literature should uncover some answers to these questions.

Since most of the significant work in the field of industrial testing completed prior to about 1935 has been adequately summarized by others, notably Bingham (5), Viteles (50), and Garrett and Schneck (22), it is not necessary to include here reviews of work completed prior to that date. Obviously, even when that mass of data is eliminated from review, the volume of literature on all phases of industrial testing since 1935 is too vast for the present review; hence consideration is limited to the use of manipulative tests in business and industry, which area is probably representative of the entire field. In addition, the references reviewed are limited almost entirely to those which included objective validity data and, with the exception of a few reports of British investigations, only work done in this country is included. Certain related, general reviews and bibliographies of industrial testing by Tiffin (45), Sells (42), Zerga (52), and Benjamin (2), are recommended to those interested in a detailed coverage of the topic although none of them is included herein.

EXTENT OF USE

To substantiate the suggestion in the second paragraph above that the use of psychological testing in industry is not nearly as extensive as it might be, certain reports can be noted.

In commenting about Link's statements mentioned previously, Taylor (43) noted in 1940 that the belief that the use of psychological

testing in industry is bound to increase, has been hopefully voiced from time to time throughout the intervening period of time since Link's article was published. He noted also that Hay (30) in the early part of 1940 listed eight companies which to his knowledge were making systematic and extensive use of tests in the selection of employees. Taylor then commented, "Granting that possibly two or three times as many other companies which did not happen to come to Mr. Hay's attention may have test programs, the number carrying on such activities certainly does not represent any sizable proportion of American industry." He suggested three reasons for the slow growth of testing programs: over-enthusiasm on the part of exponents of testing, over-skepticism by those unfamiliar with tests, and the expenditure of time and money required for the development of adequate industrial tests.

It should be noted that Taylor did not include in his commentary a part of Hay's remarks which said, "Many more companies have partial programs or are in process of developing a program, and it is probable that more serious attention is being given to psychological tests at the present time than ever before. Nevertheless it is remarkable, considering the success achieved by these companies, that every company in the country does not use psychological tests as a matter of course."

Some conflicting, but more convincing, testimony was given by Borow (10) when he reported that a National Industrial Conference Board survey made in 1936 found that slightly more than seven per cent of 2,412 firms used psychological tests. Although this is a small percentage of the total, the number is substantially larger than Hay or Taylor reported. Borow suggested in the same report that the number of companies using industrial tests in 1944 would dwarf the number listed in the National Industrial Conference Board survey, although no accurate figures were available to support his belief.

Judging on the basis of these few reports, there is no doubt that something is amiss in the use of psychological tests in industry, and that an explanation for this condition, which will at the same time suggest remedial measures, is needed.

SELECTION AND PLACEMENT

Most of the applications of testing in industry have been made in connection with the selection and placement of workers, probably because the most obvious use of tests is for those purposes, and because of the fact that management and workers can be more readily convinced of the value of tests for selection and placement than for other purposes.

In discussing occupational differences in manipulative abilities,

Teegarden (44) said, "A battery of manipulative tests makes possible rating of job applicants on specific traits of manipulative performance; speed, accuracy, delicacy of control, single or two-handed manipulation, ability to follow demonstration or instruction, to react to details, organize, maintain or increase speed, solve problems, etc."

He also reported that test performances of adults in 14 men's and 16 women's occupations show that no two occupations present identical combinations of score levels on the tests used, and mentioned the obvious conclusion that analysis of an applicant's test scores indicates the type of occupations for which he seems best fitted.

In their *Summary of Manual and Mechanical Ability Tests*, which includes objective descriptions of available mechanical aptitude tests and brief summaries of their application to various types of selection situations, Bennett and Cruikshank (3) wrote, "The assembly test or apparatus test which may be a miniature replica of the job situation, is probably a very adequate test for job selection, particularly for semi-skilled workers. No doubt much more adequate measures can be devised in this form of test."

It would be impractical to attempt to consider the use of each type of manipulative test separately, so a more logical division of test applications into occupational groups has been made.

Machine Operators. According to Hurt (32), in October of 1937 a garment factory in Marion, Virginia, started using four tests: Strength of grip, speed of arm movement determined by the length of time required to transfer 12 thimbles from one set of pegs to another and back again using one hand only, hand steadiness, which was scored by determining in how small an opening in a metal plate the subject could insert a stylus without making contact, and a general dexterity test. The latter test was considered to be the most valuable. It consisted of an inclined board with three equally spaced sockets in line at the front of the board with a counting device behind each socket. The subject punched the stylus successively in each hole as rapidly as possible for one minute. The average score was 35 and it was found that a score of 38 or better adequately predicted success of sewing machine operators. Although no quantitative validity data are presented, the author states that a few operators that were hired experimentally from the disqualified group demonstrated that they were not fitted for the job.

Using a miniature punch press in testing 25 experienced punch-press operators, Tiffin and Greenly (47) found that speed on the test is closely related to both speed and accuracy ratings made by foremen, the correlations being $-.55$ and $.63$ respectively. A high relationship

between accuracy on the test and a safety rating was also found to be present with a correlation of .60 reported.

Bennett and Fear (4) administered a series of tests to machine tool operators working on turret lathes, Bullard Automatics, precision grinders and milling machines at the plant of Martin and Schwartz, Inc. in Salisbury, Maryland. They used supervisor's ratings as a criterion and obtained a correlation of .40 with performance on the Wisconsin miniature test for the engine lathe, .41 with a hand-eye coordination test consisting of a dummy drill press, and .46 with a hand-tool dexterity test involving the use of a wrench and screwdriver. Another approach to demonstrate the usefulness of the tests was taken by the authors in reporting that 91% of the men rated excellent received excellent or good total test scores, 75% of the men rated good received scores of excellent or good, and none of the men rated below average or poor received a score of excellent.

A type of job-sample test used with sewing machine operators was reported by Blum (7). In taking the test, the applicant operated a sewing machine performing two tasks: (1) following an irregular line and (2) sewing between two parallel though abruptly irregular lines. A correlation of $-.42$ between time required to complete the test and production records was obtained. He proposed certain critical scores for further use. If a critical score of 260 were used, 20% of the poor group and none of the good group would be eliminated; with a score of 240, 28% of the poor group and 4% of the good would be eliminated; score of 220, 56% of poor and 20% of good would be eliminated; and with 200, 70% of poor and 28% of good would be eliminated.

An extension of the use of the Minnesota Rate of Manipulation Test is reported by Jurgensen (33). The Ziegler revision of the test was administered to a group of men hired to be converting machine operators in a paper mill. The performance level on this test depends upon speed of gross hand and arm movements. The men were divided into five groups by each of three supervisors on the basis of speed of work. The ratings thus obtained for each man were transformed into T-scores and the sum of the three scores for each man was used as a criterion. The maximum multiple correlation obtained was .60 for a group of 60 men.

Ross (38) found that when a critical score of 304 seconds on the O'Connor Finger Dexterity Test was used with a group of 41 machine-tool trainees, 90% of those rated "A," 100% of the "B's," 85% of the "C's," and none of the "D's" or "E's" would have been selected.

Andrews (1) reported the results of a year's experience with selection tests for engineering operatives. A battery of seven tests including in-

telligence tests and five apparatus tests was administered to 122 operators. Specifically, the apparatus tests included were a steadiness test, a finger dexterity test wherein small metal discs are placed as quickly as possible in recesses in a metal plate, a second dexterity test wherein pairs of screws under slight tension have to be unfastened and then fastened again, a bi-manual test involving simple coordinated movements of both hands, and a block sorting test made up of 100 blocks, thirty-six of which have part of a pattern design missing. It was found that if the highest scoring half of the applicant group had been hired, 87% would have been satisfactory, rather than the 60% which were satisfactory when selected without the use of tests.

Assembly Workers. Drake (16) used three tests, a pin board, a controlled turning test and a right-left turning test in experimental work for the Eagle Pencil Company. He found that those who scored above average on the test began training for a two-hand assembly job at 80% of normal worker efficiency and after two weeks training reached 97% efficiency. Those who scored below average on the tests began at 76% and ended at 91% of normal efficiency.

In a small group of watch factory workers Candee and Blum (11) found a correlation of only .26 between foremen's ratings and O'Connor Finger Dexterity Test scores. There was a critical ratio of 2.18 between mean scores on the test made by superior and mediocre workers.

Experience with employee selection tests for electrical fixture assemblers is reported by Tiffin and Greenly (46). In a group of 36 workers who performed operations of burning, twisting, and soldering ends of insulated wire a correlation of .63 was found between foremen's ratings of quality and scores on a hand precision test, and of $-.16$ between test scores and production records. They also reported a correlation of zero between ratings and finger dexterity test scores. For a group of 33 workers whose job consisted of placing plug and socket, soldering tips on ends of wires, fastening the wires and assembling the plug and socket the following relationships were reported. A correlation of .22 was found between production records and dexterity test scores (with experience held constant), and of .49 with hand precision test scores. Using supervisor's ratings of general efficiency as a criterion, a validity of .33 was obtained for the dexterity test (with experience held constant), and of .42 for the hand precision test. For a group of 44 operators whose work consisted of placing parts in a chassis and then connecting and soldering wires, the highest correlation found was only .27, that correlation being between dexterity scores and supervisor's ratings of efficiency (with experience held constant).

Blum (6) investigated the possibility of using the O'Connor Finger and Tweezer Dexterity Tests for selection of watch factory workers doing miniature assembly work. In a group of 137 workers, he found a correlation of .26 between finger dexterity time score and a salary ratio, .32 between tweezer dexterity time score and salary ratio, and .39 between a combined finger-tweezer dexterity score and salary ratio. He also reported a significant difference between time scores made by workers with less than seven days of experience and scores of workers with more than one year of experience, the latter group taking less time as might be expected.

Inspectors. Henshaw (31) used three performance tests with a group of 18 female paper inspectors. These girls inspected sheets of paper approximately 20×30 inches in size for weight and appearance by flipping the sheets over from one stack to another. All of the girls were 25 years of age or under and had at least six months of experience on the job. Using output figures as a criterion (reliability of .69 to .86) he obtained a correlation of .70 with a hand-arm dexterity test in which the subject continuously tapped two 3" keys located one foot apart one after the other. A validity of .61 was found for a tactile discrimination test in which the subject sorted a stack including three grades of paper. Using a choice-reaction test in which the subject reacted to one of three stimulus lights by pressing the correct one of three buttons to extinguish the light, a correlation of .52 was found with output records.

In studies of tests for the selection of inspector-packers, Ghiselli (23, 24) found the following correlations between a combination of ratings of supervisor and forelady: With a pegboard, $-.50$; with the Minnesota Rate of Turning Test, $-.40$; and the Minnesota Placing Test, $-.24$. The duties of the group of 26 girls consisted of (1) filling capsules, vials, and bottles with serums and antitoxins, (2) examining the filled containers for the presence of extraneous foreign matter, (3) labeling the containers, and (4) cartoning and packaging them.

Other Industrial Workers. The validity of certain mechanical ability tests for selecting cotton mill machine fixers is reported by Harrell (27). A tetrachoric correlation of .42 was obtained between scores made by 45 loom fixers on a 15 minute adaption of the Stenquist Mechanical Assembly Test and proficiency ratings made by supervisors. A correlation of .84 between scores made on the same test by 40 fixers in a carding department and ratings of mechanical ability made by an overseer was reported. The correlation between a composite of three ratings of mechanical ability and test scores of 10 spinning frame fixers was found to be .78. It is interesting to note the substantial differences between

the reported correlations when the criterion used was a rating of mechanical ability and when it was a rating of general job proficiency.

Drake and Oleen (20) reported the development and use of eight apparatus tests and one paper test in connection with studies for the Eagle Pencil Company. A comparison of the earnings of four groups differentiated on the basis of test scores is of interest. A group of 13 old employees with 32 months of experience who passed the test received wages in the amount of 116.1% of the average. A group of 10 employees with 24 months of experience who failed the tests reached a wage level of 93.6% of the average. A group of sixteen new employees who passed the tests and had only two months experience reached 97.4% of average level. A group of nine workers who passed the tests and received two weeks of training reached a level of 113.3% of average. They found that the cost of giving these tests was about two dollars per person which is less than half the difference in weekly wages between those passing the tests and those failing them.

A battery of tests including tests of motor coordination, intelligence, and emotional attitude was administered to 2,246 applicants for semi-skilled industrial jobs according to Cleeton (12). From this number, 849 applicants were selected for training, of which 546 successfully completed training. The only markedly significant difference between median test scores for the group which completed training and the group which was disqualified for apparent lack of ability was found for the motor coordination test. The median score for the first group was 75.5 and for the second, 61.3.

The use of certain undescribed machine and mechanical ability tests in the aviation industry is reported by Schultz (39). Of a group of 17 men rated "first class" by foremen, 88% received satisfactory test scores. In a group of 50 men rated as "semi-skilled," 88% also received satisfactory test scores. Only 42% of a group of 31 men rated as "bench-hands" received satisfactory scores. A year later 25 of 27 men selected by tests were doing good or average work.

Evans (21) found the median validity coefficient of a two-board, two-hand peg transfer test to be .34. The test was administered to 15 samples employed at work involving the use of two hands.

Cook (13) developed several tests on the basis of job analysis for the Western Electric Company. A coil winder test involving the task of winding wires in a prescribed manner around screws on a board differentiated a below average earning group from an above average group. In the above average group, 8% failed the test while in the below average group 72% failed. With relay adjustors, he used a monotony

test in which the worker tapped through a hole to hit a lower plate without hitting the sides of the hole in the top plate. In this group only 9% of those above average in efficiency were below average on the test, while 75% of those below average in efficiency were also below average on the test.

Moore (37) reported the use of manipulative tests in several situations. A Detroit company used the Detroit Assembly Test which consists of two fairly large boxes of 80 one-inch wooden cubes and requires the subject to use both hands to transfer the cubes from one box to the other at the same time packing the cubes neatly and in an orderly manner in the second box. When used in the selection of candy wrappers and packers, the company found that since the introduction of this test, 96% of the workers were satisfactory as compared to 52% before the test was used. He reported the use by the United States Employment Service of a wooden modification of the O'Connor Finger Dexterity Test in which the subject is required to move two pegs at a time, one in each hand, to a second board. A correlation of .45 was found between scores on this test and production records of 43 can packers. Moore also reported results of studies made by the United States Employment Service of several occupations using the Minnesota Spatial Relations Test. The following correlations were obtained between the test scores and production records or other factors: for can packers, from .20 to .24; for coding clerks, from .26 to .38; for calculating operators, from .32 to .59; for card punch machine operators, from .31 to .55; and for power sewing machine operators, from .40 to .50.

As reported in *Factory Management* (53), the Ford Motor Company used a hand dexterity test, a finger dexterity test, and a test wherein the subject is required to determine the size of rivets by touch rather than by sight. Foremen agreed that workers selected by the use of tests were superior to those selected by regular methods. In the same publication (54) it was reported that the Woodward Governor Company of Rockford, Illinois started using two tests of machine skill in 1938, and in 1943 the management stated that they were convinced that the tests are of inestimable value to the organization, not only for their value in placement but in upgrading. In one of the tests described, the subject controls a pencil by two cranks which turn in opposite directions and at different speeds and is required to follow a pattern on paper. In the other test described, the subject controls the path of a pencil by two cranks, following lines on a revolving drum. Although the two reports in this paragraph contain only subjective validity data, the tests and use made of them are especially interesting.

Crissey (14, 15) reported the use of tests in the AC Spark Plug Division of General Motors Corporation. A group of job-setters were given a number of paper and pencil tests, several well known apparatus tests, and two peg-board tests to check uni-lateral and bi-lateral dexterity. A validity of .65 for a selected battery is reported with supervisor's ratings used as a criterion. It was further reported that 69% of those who received test scores in the high third were also rated in the high third by supervisors, while none of those rated in the low third received scores in the high third of the group. A multiple correlation of .74 was found between a similar test battery and production records for aircraft spark plug gappers. The tests indicated that the most productive workers were above the average in bi-manual coordination and visual perception. In the same group, 88% of those who scored in the high third on the tests were also in the high third based on production records and all of those who had production records in the low third also received test scores in the low third. The tests also were demonstrated to be of value in the reduction of worker turnover rate when it was found that turnover for personal reasons of male employees for a one year period in a group pre-selected by tests was 5% while in a group not pre-selected the rate was 12%. He also presented one of the most convincing bits of evidence concerning the value of tests yet reported when he stated that as a consequence of hiring only those scoring in the high third on the battery of tests, the average production per operator was higher than the best previous individual record and there was considerable improvement in worker morale on the job.

As reported in *Occupations* (55), a battery of sixteen tests was administered to 51 aircraft riveter trainees. Those trainees whose total test scores were in the upper third of the total group correctly set 26% more rivets within a set period of time than those whose total test scores were in the lower third of the total group. The probability that such a difference would occur by chance is less than 1 in 1,000. Using the same timed work sample as a criterion, a multiple correlation of .60 was obtained for a battery of three tests which included a Worker-Analysis Pegboard Apparatus Part I, a Worker-Analysis Finger Dexterity Test Part II, and a figure-copying test.

Department Store Workers. Blum and Candee (8) investigated the possibility of using the Minnesota Placing Test, the Minnesota Turning Test and the O'Connor Finger Dexterity Test for selection of department store packers and wrappers. A multiple correlation of .38 was found between production records of a group of fifty-two seasonal employees and test scores while a similar correlation for permanent

workers was only .24. Efficiency on the job seemed to be the result of experience which was borne out by the fact that experienced workers scored significantly higher than seasonal workers, and that no difference in test scores was found between satisfactory and less satisfactory permanent employees. The authors suggested that these tests should not be used in selecting permanent employees.

In a later study, the same authors (9) found that manual dexterity as measured by the O'Connor Finger Dexterity and the Ziegler Placing Tests is not a selective factor for department store wrappers. With packers, a slight difference was found between the scores of experienced and inexperienced workers but this difference disappears with experience.

Ghiselli (25) found in checking the results reported by Blum and Candee that two of the same three tests plus a finger dexterity test would not be satisfactory as selection devices for package wrappers. In a group of 42 seasonal wrappers, the correlation of supervisor ratings with Minnesota Placing Test scores was $-.10$, with Minnesota Turning Test Scores $-.02$, and with the finger dexterity test scores $.02$. He suggested that if motor and dexterity tests are to be used as selective devices for this type of worker, they should be different in nature from the three used.

ACCIDENT PRONENESS

In beginning a study on the prediction and control of accidents, Drake (18) found that only moderate relationships had been reported between accident records and test scores when the tests had been handled in the conventional manner. He stated, "Far more significant relationships were found when the differences between scores on perceptual and motor tests were compared with an index that took account of both frequency and severity of accidents." The results obtained lead him to propose the hypothesis that accident proneness is a phenomenon associated with discrepancies in level between perception and motor reaction. It was observed that persons whose perceptual level is equal to or higher than their motor level are relatively safe, while those whose perceptual level is lower than their motor level are accident prone, with records of more frequent and more severe accidents than the former group. According to Drake, this implies that those who can see faster than they can react are relatively safe, while those who react faster than they can see are accident-prone. It is apparently possible however for the subjects studied to have a variety of uncorrected defects of vision and still get perceptual clues that were

quite adequate for effective and safe behavior. Drake noted that in order to make comparisons between perceptual and motor levels it was necessary to use tests that yield measures of the function intended and that are as free as possible from other complicating factors. In support of his contention that there is a marked tendency for difference scores to pick out workers with the high accident frequencies and severities, Drake offered as substantiating data, a comparison between accident indexes and difference scores. It is sufficient here to note that the half of a group of industrial workers that had the best difference scores had an accident index of 7.71 while the low half had an accident index of 25.04. He also reported that in one new group of employees selected on the basis of test and difference scores, accident reduction was 70% of the rate of workers selected by previous methods.

Williams (51) reported the use of a series of 15 simple tests in a study of accident prevention. The series included an interrupted pursuit test, a hand-eye coordination test involving dotting of small circles on a revolving disk moving at an increasing rate of speed, reaction to visual or auditory-tactual stimulus by pressing correct button, muscular hand-arm steadiness, and a hand dynamometer. Those scoring in the lowest quartile had an average accident history of .73 accidents per man per year while the average accident rate of the remaining 75% was only .37 per man per year.

TRAINING

An interesting approach to the value of testing devices in training programs is reported by Tiffin and Lawshe (48). They found that hosiery mill employees with the poorest finger dexterity as measured by the Purdue Peg-Board cost a company \$59 each in minimum make-up before they made the rate, while employees having the best dexterity cost the company only \$36.40. They suggest that tests can answer three questions about training: (1). Who should be trained? (2). Where should training begin? (3). Has training been adequate?

Martin (36) reported the use of a battery of written and apparatus tests by the Woodward Governor Company. He stated that the company hadn't lost a member of the organization because of technical inability during the year since the tests were begun and that as a measure for weeding out the untrainable, the tests had worked out just about 85%.

Knowles (34) reported the experience of Northwest Airlines in using a battery of tests with general mechanics. The battery included one written test designed to measure ability to learn, a carefulness test in

which the subject is given a pile of metal pieces to assemble into units exactly like models, and a simple assembly test in which the subject assembles rows of bolts in a plate alternately from front and back while the plate is shielded from his view. In a group of 100 men hired for training, 25 were rated good, 50 fair, and 25 poor. If the 50 scoring highest on the test had been chosen, 38% would be in the good group, 58% fair, and only 4% poor. If the top 25 had been hired, 56% would have been in the good group, 44% fair, and none poor.

DEVELOPMENT OF TEST PROGRAMS

One of the striking impressions gained through reading published reports of industrial application of tests is the lack of ingenuity evidenced by many, if not most, investigators. The most successful test programs that have been reported included tests that were designed especially for specific situations. As could readily be discerned, however, the common procedure in too many instances was to get together a large group of tests that had worked in other places and administer them, hoping that some would prove to be of value. It also seemed that many times tests were administered and then the most expedient criterion was selected. Not disregarding certain advantages of standardized and ready-made tests, it seems painfully obvious that low validities in many situations were due to "shot-gun" test applications and careless selection of criteria. The resultant apparent limited value of many testing programs doesn't prove that tests are useless, but patently suggests that new tests should be developed and that any standardized tests used should be selected carefully on the basis of job analysis and possibly factor analyses, and that criteria should be chosen just as carefully.

Improvement of Criteria. In reviewing the literature the authors simply gained support for their own contention that one of the major problems in the effective use of tests in industry is the difficulty in obtaining adequate criteria. About eighty percent of the articles available for review were only generally descriptive and included no acceptable validity data. When given, validity data were often expressed in subjective terms, the usefulness thus being limited. Since all validation of tests, whether the methods used are simple or statistically complicated, depends on the availability and selection of adequate criteria, the first step in any development program must be the isolation of suitable criteria. To be most effective, a criterion must be reliable, relevant, and free from bias. In order to obtain criteria that meet these conditions, it is quite probable that new record systems of

production, employee ratings, absenteeism, accidents, tenure, training, and the like will have to be installed in many companies.

Drake (17, 19) suggested that all conventional criteria be discarded and proposed that test results be used as the criterion against which to measure the success of management, particularly the success of supervisors in releasing the measured abilities of workers in production. Apparently he found himself forced to make this proposal when he found that foremen's rankings, output figures, and efficiency ratings by time study techniques were proved to be unreliable as measures of operator ability on the job. This break with generally accepted procedure apparently ignores such considerations as motivation which would certainly be operating if the applicant felt he was setting his own performance level at the time of testing. Furthermore, a reliable criterion is only half of the solution since validity is at least equally as important as reliability.

Job Analysis. The necessity for the use of job analysis in developing a testing program for selection of mechanical workers was pointed out by Moore (37) when he wrote:

The level and type of mechanical ability that is needed depends upon job demands, and varies from one type of job to another. An assumption that is frequently made, and that has been as often disproved, is that there is such a unit as general mechanical ability. Mechanical ability can only be appraised when the job demands have been analyzed, interpreted, and converted into bodily elements. Therefore the selection of tests with which experimental work is to be carried on is determined by a careful analysis into human demands, a grouping of these demands in terms of functional similarities, and a preliminary selection of those tests which have been found valuable in appraising the same functions in similar situations.

Drake has written a number of articles in which he advocates the use of job analysis as the best possible basis for test development. In early studies reported with Oleen (20) the following job factors were considered useful in designing tests:

1. Length of cycle,
2. Nature of the elements of the cycle,
3. Sizes of materials and parts,
4. Serial order of elements of the cycle,
5. Three-dimensional position of parts manipulated,
6. Incidence of finger, wrist, arm, and body movements,
7. Posture of the operator,
8. Visual, tactual, and kinesthetic attentive factors,
9. Speed and rhythm of work.

When working with the Johnson and Johnson Company at New

Brunswick, New Jersey, to find suitable ways to measure observed differences in ability used on the job, Drake (17) found reason to comment, "We employed the methods then common to industrial psychology and still, I regret to say, too common in this field. We applied batteries of paper and pencil tests . . . and tried to find significant correlations between tests and various measures of success on the job such as foremen's ranks, production percentage efficiency, etc. This older procedure was cumbersome and costly and was characterized by trial and error to the extent that made it seem unscientific and illogical in the extreme." From the techniques developed in this study using time study analyses of jobs and groups of related jobs, six types of measurable human abilities were isolated. They were:

1. General finger, hand, wrist, arm dexterity,
2. Dual hand dexterity,
3. Bilateral hand dexterity,
4. Hand and foot coordination,
5. Machine tending ability,
6. Inspection ability.

Several performance tests were designed which had obvious similarities with the jobs studied and Drake reported that when these tests were administered again to the same operators after a lapse of time, sufficiently similar scores were obtained to indicate that the tests measured something relatively unchanging.

Cook (13) reported that he did not use detailed job analyses in selecting tests for his first work with the Western Electric Company, but followed the procedure of administering as many as 14 or 15 tests to the workers. He stated that the usual results were that two or three of the tests would prove to be of selective value. This procedure was considered to be wasteful of time and effort and consequently other procedures were investigated. Ultimately job analyses were developed which included nine factors as a basis for test selection. The factors included were intelligence, eye and hand coordination, finger dexterity, manual dexterity, small-tool dexterity, repetitiveness of work, accuracy, range of observation, and visual memory. Cook wrote that these factors were selected because it was felt that adequate tests were available for measuring an applicant's performance on any and all of them. Lest it be overlooked, it should be pointed out that the selection of job analysis factors on the basis of available tests makes use of only a portion of the potential value of the job analysis results to a test development program.

It seems probable that many psychologists working in industry who

have attempted to use job analyses in test development have not used them in the best possible manner. There has been too much of a tendency to designate abilities differentiated on the basis of superficial descriptions of the job or of worker abilities required to perform efficiently on the job. More thorough job analyses including the use of motion study to identify necessary bodily performance skills, and the identification of significant psychological traits and abilities such as visualization, numerical facility, and perceptual speed, should afford better bases for the selection and development of tests.

Factor Analysis. The use of factor analysis as a means of developing tests for industry is a forward step still to be used to the best possible advantage. Even this method cannot be utilized to its full potentiality, however, until more and better tests are included in the analyses.

Harrell (28, 29) reported a factor analysis of mechanical ability tests which included thirty variables, eleven of them apparatus tests including three pinboards, a pegboard, a peg sorting test, two tests of manipulating nuts and bolts, Crocket's block packing test, a test involving the placing of blocks along strips, a nut, bolt, and screwdriver manipulative test, and the wiggly blocks test. The tests were administered to a group of ninety-one cotton mill machine fixers. He found five factors which he named perception of detail (P), verbal relations (V), visualizing spatial relations (S), manual agility (A), and youth (Y). The tests which appeared in the agility factor were pinboard with both hands, two of the nut and bolt manipulation tests, block packing, pinboard with non-preferred hand, wiggly blocks, and a paper and pencil dotting test. Harrell noted that the P, V, and S factors have been previously identified by Thurstone and others. He suggested that since certain of the group paper and pencil tests measure each of the factors present in the manual tests, other paper tests can probably be devised to replace all of the manipulative tests. Considering the list of tests included by Harrell in the analysis, it seems highly improbable that much attention was paid to job analysis and that no attempt was made to develop tests especially to measure the abilities or aptitudes needed by machine fixers. He probably is correct when he suggests that paper and pencil tests could be devised to measure the abilities measured by manipulative tests that were included in the battery, but it is quite possible that better and more apropos manipulative tests could be developed that paper tests could not easily replace. It would also seem appropriate to question the necessity for the inclusion of tests for the measurement of mechanical abilities which are highly weighted with a

verbal factor, if equally valid manipulative tests could be developed that exclude a verbal factor.

Seashore (40) found that motor skill tests isolated into groups by factor analysis usually involved a similar pattern of movement regardless of musculature or sense-field required in subject performance which is a functional rather than an anatomical grouping of the skills. Later, Seashore in conjunction with Buxton and McCollum (41) suggested that the next step in identifying statistically located factors would be to use the technique of photographic motion study and simple introspection by the subjects rather than the all-too-frequent second hand analysis or vicarious introspection of the experimenters. They further suggested that any one physiological limit, such as speed or muscle contraction, is usually important only for a given work method and that changing to another work method may either partially or entirely overcome this limitation. This suggestion serves as one explanation of the success of applications of specially designed tests and the relative lack of success when shot-gun application of tests is made.

Guilford (26) makes some valuable suggestions in pointing out the results of certain factorial studies. He wrote:

Buxton's recent preliminary analysis tends to show that there are common motor factors. As Seashore points out, however, intercorrelations among motor tests are notoriously low and clusters are limited in scope. This fact forecasts a number of motor abilities of narrow range and tests with relatively low communalities. General factors of broad scope among motor abilities may include physical strength, agility, and steadiness of control.

The application of factor analysis to test development in industry is still rare, probably because the pressure to produce something immediately applicable does not often permit the use of a relatively long range program of factor analyses involving an accumulation of the necessary amount of intercorrelational data regarding tests and criteria. With sufficient data, the factorial content of a criterion can be determined and thus the criterion becomes a much more meaningful target for test development. One of the more obvious uses of the factorial approach to test development is to attempt to design pure tests to measure those unique factors that are demonstrated to be present in the criterion. As another result of factor analysis, the value of each test in predicting the known factors can be determined, thus permitting a close estimate of the validity of a test and the avoidance of the possibility of including more than one test in a battery which actually measure substantially the same thing. It is also possible that a particular test may be substantially loaded with a factor found to be related to success on a

job even though the test task has little apparent relation to job performance. The value of such a test probably would not be fully realized unless made evident as the result of factor analysis. It thus seems clear that factor analysis can assist considerably in making an industrial test research and development program effective.

CONCLUSIONS

When using published material as a basis for judging the effectiveness of the use of psychological tests in business and industry, it must be remembered that many companies do not publish the results of their testing programs and it is quite probable that much successful and unsuccessful work remains unknown to all but a few. Nevertheless, a consideration of the published accounts of work done with manipulative tests would seem to indicate several conclusions in answer to the questions raised in the introduction of this review.

1. There is no doubt that testing programs can be most effective when the psychologists responsible for those programs are thoroughly familiar with industrial problems.

2. In many instances it may be necessary to redesign employee rating and record keeping systems to afford adequate criteria.

3. The effective combination and use of correct job and factor analyses in test selection and development will contribute greatly to the rapid development of extensive and effective psychological testing programs in industry.

4. The practice of reporting test results in other than technical terminology will probably contribute to the more ready acceptance of psychological testing programs by industrial management and workers.

5. Some of the most valuable applications of tests in industry may well be for purposes other than selection and placement of workers.

If the above suggestions are incorporated into psychological testing programs, there should be a much more rapid expansion of effective programs in business and industry than has been apparent to date.

BIBLIOGRAPHY

1. ANDREWS, A. C. A year's experience of selection tests for engineering operatives. *Occup. Psychol.*, 1944, **18**, 126-130.
2. BENJAMIN, H. C. Employment tests in industry and business: a selected, annotated bibliography. *Industr. Relat. Sect.*, Princeton Univ., Bibliog. Series No. 67 (Rev.). Princeton, 1945.
3. BENNETT, GEORGE K., & CRUIK-
SHANK, RUTH M. *A summary of manual and mechanical ability tests (preliminary form)*. New York: The Psychological Corporation, 1942.
4. BENNETT, G. K., & FEAR, R. A. Mechanical comprehension and dexterity. *Person. J.*, 1943, **22**, 12-17.
5. BINGHAM, W. V. *Aptitudes and aptitude testing*. New York: Harper, 1937.
6. BLUM, M. L. A contribution to man-

- ual aptitude measurement in industry: the value of certain dexterity measures for the selection of workers in a watch factory. *J. appl. Psychol.*, 1940, 24, 381-416.
7. BLUM, M. L. Selection of sewing machine operators. *J. appl. Psychol.*, 1943, 27, 35-40.
 8. BLUM, M. L., & CANDEE, BEATRICE. The selection of department store packers and wrappers with the aid of certain psychological tests, I. *J. appl. Psychol.*, 1941, 25, 76-85.
 9. BLUM, M. L., & CANDEE, BEATRICE. The selection of department store packers and wrappers with the aid of certain psychological tests, II. *J. appl. Psychol.*, 1941, 25, 291-299.
 10. BOROW, H. The growth and present status of occupational testing. *J. consult. Psychol.*, 1944, 8, 70-79.
 11. CANDEE, BEATRICE, & BLUM, M. L. Report of a study done in a watch factory. *J. appl. Psychol.*, 1937, 21, 572-582.
 12. CLEETON, G. U. Implications of industrial selection by tests. (Abstract.) *Psychol. Bull.*, 1938, 35, 678-679.
 13. COOK, D. W. Psychological tests for unskilled jobs. In *Psychological aids in the selection of workers*, Personnel Series No. 50. New York: American Management Association, 1941.
 14. CRISSEY, O. L. Tests predictive of success in occupation of job-setter. *Psychol. Bull.*, 1942, 39, 436.
 15. CRISSEY, O. L. Aptitude testing improves shop morale. *Amer. Mach.*, 1944, 88, 91-94.
 16. DRAKE, C. A. Spot the best by test; dual operations often effect time savings. *Fact. Mgmt.*, 1937, 95, 60-62.
 17. DRAKE, C. A. New developments in the selection of factory workers. In *Selection and development of foremen and workers*, Production Series No. 127. New York: American Management Association, 1940.
 18. DRAKE, C. A. The prediction and control of accidents. *Sci. Mon.*, N. Y., 1940, 51, 74-76.
 19. DRAKE, C. A. Job tests. *Person. J.*, 1941, 20, 184-189.
 20. DRAKE, C. A., & OLEEN, H. D. The technique of testing. *Fact. Mgmt.*, 1938, 96, 71-78.
 21. EVANS, S. D. A performance test for bi-manual jobs. (Abstract.) *Psychol. Bull.*, 1940, 37, 448-449.
 22. GARRETT, H. E., & SCHNECK, M. R. *Psychological tests, methods, and results*. New York: Harper, 1933.
 23. GHISELLI, E. E. Tests for the selection of inspector-packers. (Abstract.) *Psychol. Bull.*, 1941, 38, 735.
 24. GHISELLI, E. E. Tests for the selection of inspector-packers. *J. appl. Psychol.*, 1942, 26, 468-476.
 25. GHISELLI, E. E. The use of the Minnesota Rate of Manipulation and the O'Connor Finger Dexterity Tests in the selection of package wrappers. *J. appl. Psychol.*, 1943, 27, 33-34.
 26. GUILFORD, J. P. Human abilities. *Psychol. Rev.*, 1940, 47, 367-394.
 27. HARRELL, W. The validity of certain mechanical ability tests for selecting cotton mill machine fixers. *J. soc. Psychol.*, 1937, 1, 279-282.
 28. HARRELL, W. A factor analysis of mechanical ability tests. (Abstract.) *Psychol. Bull.*, 1939, 36, 524.
 29. HARRELL, W. A factor analysis of mechanical ability tests. *Psychometrika*, 1940, 5, 17-33.
 30. HAY, E. N. The use of psychological tests in selection and promotion. *Personnel*, 1940, 16, 114-123.
 31. HENSHAW, L. S. Selection tests for paper sorters. *Hum. Factor, Lond.*, 1937, 11, 146-153.
 32. HURT, J. Evaluating applicants by dexterity testing. *Employ. Serv. News*, 1939, 6, No. 6, 7-8.
 33. JURGENSEN, CLIFFORD E. Extension

- of the Minnesota Rate of Manipulation test. *J. appl. Psychol.*, 1943, 27, 164-169.
34. KNOWLES, W. We've stopped hunch hiring. *Nation's Bus.*, 1945, 33, 23-24.
 35. LINK, H. C. *Employment psychology*. New York: Macmillan, 1919.
 36. MARTIN, I. C. How and why Woodward Governor Company uses aptitude tests for employees. *Sales Mgmt.*, 1941, 48, 18-20.
 37. MOORE, H. *Experience with employment tests*. New York: National Industrial Conference Board, 1941.
 38. ROSS, L. W. Results of testing machine-tool trainees. *Person. J.*, 1943, 21, 363-367.
 39. SCHULTZ, R. S. Personnel selection in the aviation industry. *Person. J.*, 1940, 19, 99-105.
 40. SEASHORE, R. H. An experimental and theoretical analysis of fine motor skills. *Amer. J. Psychol.*, 1940, 53, 86-98.
 41. SEASHORE, R. H., BUXTON, C. E., & MCCOLLOM, I. N. Multiple factorial analyses of fine motor skills. *Amer. J. Psychol.*, 1940, 53, 251-259.
 42. SELLS, S. B. Measurement and prediction of special abilities. *Rev. Educ. Res.*, 1944, 14, 38-54.
 43. TAYLOR, H. C. A test program in an industrial company. In *Values of psychology in industrial management*, Personnel Series No. 43. New York: American Management Association, 1940.
 44. TEEGARDEN, L. Occupational differences in manipulative abilities. *Psychol. Bull.*, 1936, 33, 736.
 45. TIFFIN, J. *Industrial psychology*. New York: Prentice-Hall, 1943.
 46. TIFFIN, J., & GREENLY, R. J. Employee selection tests for electrical fixture assemblers. *J. appl. Psychol.*, 1939, 23, 240-263.
 47. TIFFIN, J., & GREENLY, R. J. Experiments in the operation of a punch press. *J. appl. Psychol.*, 1939, 23, 450-460.
 48. TIFFIN, J., & LAWSHE, C. H., JR. How tests can strengthen the training program. *Fact. Mgmt.*, 1944, 102, 119-121.
 49. VITELES, M. S. Psychology in industry. *Psychol. Bull.*, 1930, 27, 567-635.
 50. VITELES, M. S. *Industrial psychology*. New York: W. W. Norton, 1934.
 51. WILLIAMS, A. W. Accident prone workers can be weeded out by simple tests. *Textile World*, 1943, 93, 68-69.
 52. ZERGA, J. E. The development and use of apparatus tests in industry. *J. appl. Psychol.*, 1944, 28, 199-202.
 53. *How three companies use tests to select and place people*. [Anon.] *Fact. Mgmt.*, 1943, 101, 113-120.
 54. *Use the right selection and testing methods*. [Anon.] *Fact. Mgmt.*, 1943, 101, 96-105.
 55. *Ten years of occupational research*. Division of Occupational Analysis and Manning Tables, Bureau of Manpower Utilization, War Manpower Commission. *Occupations*, 1944, 22, 387-446.

THE TREATMENT OF QUALITATIVE DATA BY "SCALE ANALYSIS"

LEON FESTINGER

Research Center for Group Dynamics, Massachusetts Institute of Technology

The basic problem of measurement in connection with measuring instruments such as questionnaires and interviews is to arrange a number of individuals in rank order from high to low on the basis of their responses to a series of questions. The solution of this measurement problem, that is, being able to say that one individual is higher, lower, or equal to another individual, along some specified dimension, has been attempted in a variety of ways. Dependence upon expert ratings, examination of intercorrelations among items and correlations between individual items and total scale scores are some of the methods which have been used to facilitate scale construction.

During the war Dr. Guttman, together with others in the Research Branch of the Information and Education Division of the War Department, developed a new technique for analyzing qualitative data which they have called "scale analysis" (2, 3, 4, 12). This technique attempts to solve some of the problems of scale construction which have up to now not been adequately handled.

There have as yet been relatively few publications concerning "scale analysis." The technique has, however, received quite a bit of favorable attention from social scientists. McNemar (7), for example, in a recent article make the following statements:

The next step would be to construct a uni-dimensional scale for measuring each of several components. Use of the Guttman technique would greatly facilitate the meeting of this requirement. . . . A Guttman-type scale for each component would also avoid the absurdity of having several questions, which supposedly tap a given component turn out . . . to have little or nothing in common.

. . . we have stressed the basic need for reliability, validity, and uni-dimensionality for the instruments or devices used to classify or measure individuals with respect to their opinions or attitudes. . . . Unitary scales can be developed by the Guttman scaling technique.

It is the purpose of this paper to review the published literature and some unpublished material with reference to the theory of "scale analysis," techniques of scale construction using "scale analysis," and the evaluation and interpretation of the scales achieved by this method.

THE THEORY OF "SCALE ANALYSIS"

The meaning of uni-dimensionality. When an investigator attempts to measure a given variable (e.g. attitudes toward Russia, intelligence, weight) he assumes, at least implicitly, that he is dealing with something which has a unitary character. That is, that he can obtain measurements which relate unequivocally to this variable.

If the concept in question, or if the measuring instrument used, is not of this type, then there will unavoidably be some equivocation in the use of the measurement scale derived. To take a very simple illustration, suppose we were measuring individuals with an instrument (e.g., a series of questions) which simultaneously measured prejudice toward Negroes and mathematical ability. Two individuals who received the same score on this measuring instrument need not at all be alike. One could be more prejudiced and the other more able mathematically. Similarly, if one individual gets a higher score than another individual, we would be uncertain to what to attribute the difference. The difference might be one of degree or it might be a difference in kind.

In any measuring instrument which is not uni-dimensional, that is, which does not measure one and only one thing, the ordering of individuals by virtue of their scores does not have the simple quantitative properties which we desire in a scale. On an intelligence test, for example two individuals may achieve the identical score in very different ways. One individual might have good mathematical ability and another individual good verbal ability. These two quite different individuals nevertheless may come out with the same total score. In what sense then can it be said that the two are equal in intelligence or in what sense can it be said that a person who gets a higher score has more intelligence than a person who gets a lower score? The differences which we obtain between individuals may be differences in type of intelligence or type of ability, in addition to differences in the degree to which they possess these abilities.

The conclusion should not, of course, be drawn that the scores on intelligence tests, or the scores on any other measuring instrument, which does not possess uni-dimensionality, have no meaning. On the contrary, they can still be useful and have a good deal of meaning (1). It can be said, however, that if such a test could be split into several components, each of which did possess uni-dimensionality, then the value of this instrument would be greater than the value of the previous test.

Where it is feasible, the attempt should be made to have our measuring instruments measure only one dimension or one variable at a time.

Even if this is not possible, it is obviously desirable that the investigator have some indication of the extent to which his measuring instrument departs from the ideal of uni-dimensionality.

The determination of uni-dimensionality. The problem still remains of determining whether or not a given measuring instrument possesses uni-dimensionality. Let us examine the characteristics which such a measuring instrument would have.

It is implied by the preceding discussion that a uni-dimensional measuring instrument would be such that a given score on that instrument could be obtained only from one pattern of responses. In addition, for these scores to represent an ordering of individuals with respect to some variable there would have to be a certain type of consistency among the responses to the various items making up the measuring instrument (4, 11, 12).

Let us examine what the data would look like if secured with a measuring instrument known to be uni-dimensional. Suppose that in order to measure the height of individuals a measuring instrument were chosen which consisted of ten sticks, each one of different length. The operation of measurement is to stand each stick up alongside each individual and to record whether he is taller or shorter than the stick. We record a plus if he is taller and a minus if he is shorter. Our data for each individual then consist of ten marks, each of which may be either plus or minus.

We might now determine how many individuals received a plus on each of the measuring sticks. We could then arrange our sticks in order from the one where the greatest number of people received a plus down to that stick where the smallest number of people received a plus.

We would now find that any individual who received a minus on stick "1" would have received minuses on all of the other sticks; an individual who received a plus on stick "1" and a minus on stick "2" would have received minuses on all of the subsequent sticks and so on. The individual who had received a plus on stick "10" would have received pluses on all of the other sticks. Each individual would thus have fallen into any one of 11 "patterns of response." These patterns could be arranged in order from "tallest" to "shortest" and we would thus have a scale for measuring height. It may be noted that since this instrument was uni-dimensional, it would have been impossible for an individual, for example, to have received a minus on stick "1," a plus on stick "2," and a minus on stick "3." In fact, there are 1013 possible "patterns of response" which would *not* have occurred.

If our ten measuring sticks were ten questions which attempted to

The scalogram board is constructed so that any row or any column can be lifted and moved to another position. The procedure of analysis is then to indicate in the appropriate column and appropriate row of the board whether the individual did or did not make that particular response. Then, by shifting rows and columns one can attempt to arrange the board so as to approach as closely as possible to the ideal "parallelogram." If the data can be made to form a parallelogram one can conclude that the measuring instrument possesses unidimensionality. The scale patterns conforming to the various scale positions become apparent from the resulting diagram.

The number of scale positions is, of course, determined by the number of questions in the measuring instrument and the number of possible responses to each question. In the present example where there are ten questions each of which has two categories of response there would be 11 possible scale positions. In general, the number of possible scale types is given as follows (10): "Add unity to the total number of categories in all questions and subtract the number of questions." In other words, if our measuring instrument were made up of seven questions, one of which had two categories, four of which had three categories, and two of which had four categories, the number of possible scale types would be sixteen.

The scalogram board has been used extensively by the Information and Education Division of the War Department. There are several limitations, however, to its use.

1. A scalogram board is fairly expensive to construct.
2. The number of rows in the board must be large enough to accommodate the number of cases in the sample with which the investigator is dealing. If a board is built which will handle 100 cases it will not be adaptable to any sample larger than this. The limitation on the number of questions and number of categories in each question because of the number of columns in the scalogram board is of the same nature.
3. The method is not very rigorous. Outside of possible hunches which the investigator may have concerning how the questions will scale themselves, the method is largely one of trial and error and inspection.
4. Since by the very nature of the method, the investigator manipulates chance deviations to his advantage in trying to form a "parallelogram," an evaluation of the resultant scale is somewhat difficult to make.

The Tabulation Technique. The use of this method for "scale analysis" is described by Goodenough (2). It is probably the simplest method available. The steps are as follows:

1. One computes the number of people in the sample making each response to each question.

2. Each question is then graphically represented by a horizontal bar with proportional areas marked off indicating the number of people making each response. These bars for each question are placed directly under one another. Fig. 2 is a reproduction of such a diagram (10) based upon seven questions concerning the desire of enlisted men to return to full time school.

3. Vertical lines are then drawn through all of the bars at the points separating categories of response in each question. The responses between any pair of these vertical lines indicate one of the "consistent" patterns of response.

4. The investigator then goes back to his data to see how well the response patterns of the individuals in his sample fit these scale types.

This procedure is more rigorous than using a scalogram board. It is apparent, however, that in order to use this procedure it is necessary to know beforehand the directions of the responses to each question with regard to the dimension being measured. This means that in order to use this tabulation technique it is necessary to know, for example, that the four possible responses on Question 6 in Fig. 2 go in the direction indicated in the diagram with respect to desire to return to full time school. This is generally not too serious an obstacle.

The originators of "scale analysis" propose that if one finds that the individuals in a sample do not adequately fit the scale patterns one may then examine the data to see if some of the responses to a question may be combined to make the fit more adequate. Thus, for example, an investigator might find that by combining the categories of "undecided" and "yes" on a given question his data might better approximate the scale patterns. The limitation to this is, of course, that again the investigator takes advantage of chance variation and evaluation becomes somewhat difficult.

The Cornell Technique. This technique is so named because it has been used at Cornell University. It has been called the "trial scoring and graphic technique" by Goodenough (2). It is described in detail by Guttman (6). It consists essentially of making a trial rank order of individuals and then examining to what extent the responses show the desired type of consistency according to this rank order. Consequent rearrangements in order and combinations of categories are made so as to approximate better and better to the type of consistency expected from a uni-dimensional scale. The investigator can then evaluate to what extent the responses fit the scale patterns. The technique rests on essentially the same principles as the tabulation technique discussed above. It has relatively little to recommend it as compared to the tabulation technique. It is more cumbersome and somewhat less rigorous. Its advantage would seem to be that it preserves somewhat greater flexibility for manipulating the items.

It may be noted that this procedure has been adapted for use with IBM equipment by Noland (8, 9).

Least Squares Method. This is another method which was devised by Guttman (3). We shall do no more than mention its existence here since, as Goodenough (2) says, "This method is too laborious to be usable when one is dealing with more than a few items and categories."

CRITERIA FOR DETERMINING THE EXISTENCE OF UNI-DIMENSIONALITY

Whichever of the above procedures is used for constructing a scale, the final problem which the investigator must face is that of making a decision as to whether or not his measuring instrument possesses uni-dimensionality. In other words, how well do his data fit those conditions which would exist if uni-dimensionality were present?

For the purpose of making a judgment concerning the scalability of groups of items, the authors of "scale analysis" have introduced the concept of *reproducibility*. From the statements in the published articles, it is somewhat difficult to understand just what is meant. The following are examples of statements made about it:

Perfect scales are not found in practice. In the past, areas the component items of which were at least 85% reproducible from a rank order have been called scales. Recent work shows that it may be desirable to be more stringent about errors and to restrict the word *scale* to areas the items of which are about 90% reproducible (12).

The degree of approximation to perfection is measured by a *coefficient of reproducibility* which is the empirical relative frequency with which values of the attributes do correspond to intervals of a scale variable. In practice, 85% perfect scales or better have been used as efficient approximations to perfect scales (4).

From these statements one would conclude that the reproducibility would be measured as follows: The responses of an individual which exactly fit one of the scale patterns would be 100% reproducible from his scale score. Responses of an individual which deviated by 1 from one of the scale patterns would be $100(n-1)/n$ percent reproducible where n is the number of questions being scaled. Thus, if in a sample of a hundred individuals who had each answered five questions, fifty exactly fitted the scale patterns, forty were 1 off a scale pattern and the rest were 2 off a scale pattern, the reproducibility would be 88%.*

* Goodenough (2) tends to give a different impression of what is meant by reproducibility. One gets the impression from this article that what is meant by 85% reproducibility is 85% of the individuals falling exactly in the scale patterns. This is obviously not what is meant by other authors.

In an example of a set of items which are considered to show a high degree of scalability one reads,

There are but 16 possible scale types out of a total of 2,592 possible types. Actually almost 400 types occurred among the 2,000 men. Over one-fourth of the men were perfect scale types, and almost all the rest were one or two responses off a perfect scale type (10).

Guttman (4, 11, 12) has also made the distinction between scales and quasi-scales, the latter being measuring instruments which are less than 85% reproducible. As the subsequent discussion will show, this is entirely an arbitrary distinction. It would, perhaps, be better not to speak of uni-dimensional scales (except in very rare instances) and to content oneself with a description of the extent to which the scale on hand departs from the ideal of uni-dimensionality.

Let us now examine how adequate a criterion 85% or 90% reproducibility is for deciding whether or not only one dimension is present. Let us take an example of five questions, each of which requires a true or false answer. We are to determine whether or not these five questions can be scaled, that is, whether uni-dimensionality is present. To make the example more specific let us suppose that to Question 1, 80% of the sample answered *false*; to Question 2, 60% responded *false*; to Question 3, 50% responded *false*; to Question 4, 40% responded *false*; and to Question 5, 20% responded *false*. The number of possible response patterns to these five questions is 32. Only six of these, however, would be "scale patterns." If *false* indicated the same direction of information or attitude on all questions, these six scale patterns would be (1) FFFFF, (2) FFFFT, (3) FFFTT, (4) FFTTT, (5) FT TTT, (6) TTTT T. If uni-dimensionality exists only these six would appear in the data.

From this information alone, namely, the fact that there are six scale patterns out of 32 possible patterns, one could not very well estimate the probabilities of obtaining these six patterns. Let us calculate the chance probability of occurrence of the six scale patterns (holding marginal totals constant), assuming complete independence among the questions. Since 80% answered *false* on Question 1, let us consider that "F" on Question 1 has a probability of .8 and that "T" on Question 1 has a probability of .2. Let us consider the probabilities of false and true responses on other questions in a similar manner. We then come to the conclusion that by chance 42.2% of the individuals would fit the scale patterns exactly. We also come to the conclusion that 48.4% of the individuals would only be 1 off a scale pattern and 9.4% would be 2 off a scale pattern. The reproducibility that one would

get by chance would be 86%. It is clear then that for this specific example a criterion of 85% reproducibility would be a very poor one to employ.

Let us examine what the chance reproducibility would be if we had nine such True-False items that we were to make into a scale. Let us assume that the proportion of "F" answers to the questions are in order .9, .8, .7, .6, .5, .4, .3, .2, .1. There would now be 2^9 or 512 possible response patterns. Only 10 of these would be *scale* patterns. One finds, however, that 18.5% of the individuals would by chance exactly fit a scale pattern and that 42.3% of the individuals would by chance be only 1 off a scale pattern. In fact, one finds that by chance one would expect about 83% reproducibility. Again it is apparent that the criterion of 85% or 90% reproducibility would be a very poor one to apply for determining the presence of uni-dimensionality.

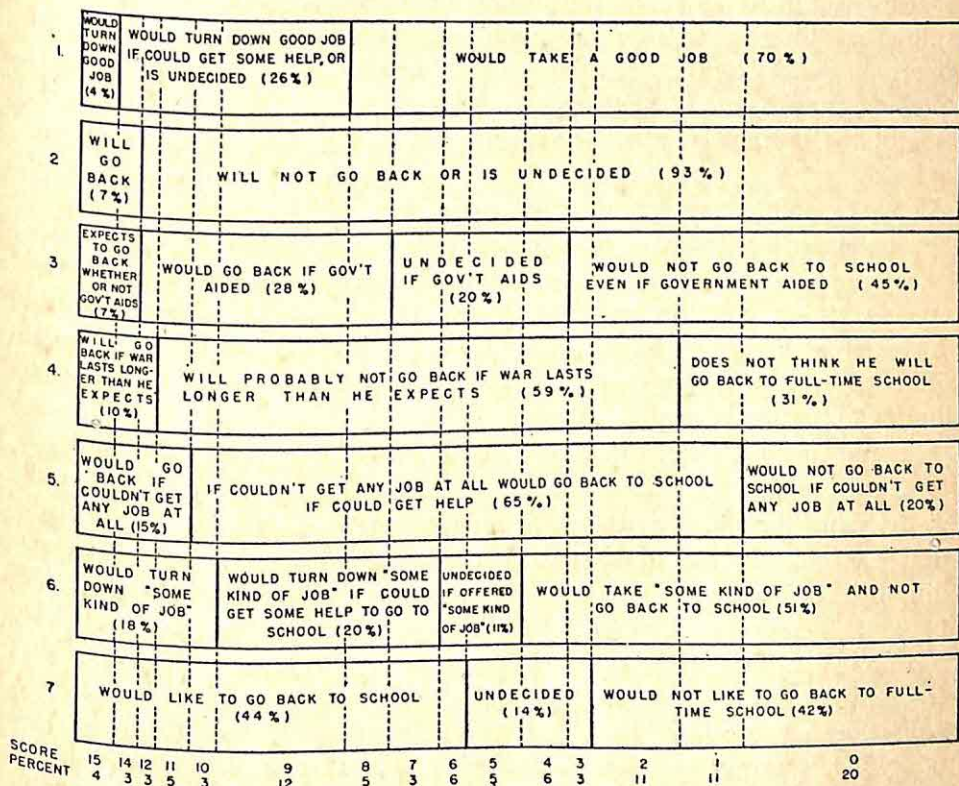


FIG. 2. SCALE OF DESIRE OF ENLISTED MEN TO RETURN TO FULL TIME SCHOOL

It is clear that applying a criterion like 85% or 90% reproducibility to all attempts at scaling, irrespective of the number of items involved or the number of possible answers to each item, leads to false conclusions. In one case where there are many items and many parts to each question, 85% reproducibility might be excellent consistency; in another case 85% reproducibility might represent no better than chance occurrence and be no evidence at all for uni-dimensionality. Any criterion for deciding whether or not there is uni-dimensionality must rest upon calculations based on the number of questions, the number of parts to each question, and the percentages of the sample in each part to each question.

Let us take an example of an actual scale (10, see Fig. 2). Seven items were used. The first item had three possible responses, the second two possible responses, the third had four possible responses, the fourth and fifth items had three possible responses each, the sixth item had four possible responses and the seventh item had three possible responses. There are 16 *scale* patterns. The total number of possible patterns is 2,592. The author states that over one-fourth of the men tested showed perfect scale patterns. He does not present the reproducibility for the scale. Let us see what evaluation we can make of the evidence for or against uni-dimensionality. One can calculate that by chance 7% of the cases would fall into perfect scale types. If we compare this theoretical 7% with the obtained approximate 25% one would certainly come to the conclusion that more people fell into perfect scale types than one would expect by chance in a sample of 2,000 cases. The major question is, however, whether or not this is evidence for uni-dimensionality. Let us assume that there were two dimensions operating. One could still obtain a greater proportion of perfect scale types than one would expect by chance.

Suppose we attack the problem from the other end. Suppose we reason that the unreliability of the responses to the questions is what makes for deviation from the perfect scale pattern when uni-dimensionality does exist. If we assume that for each of the questions 90% of the responses are "reliable" responses, (this would be equivalent to about 80% agreement on test-retest) we would conclude that 47.8% would fall into the correct scale patterns and that 52.2% would not do so because of unreliability with uni-dimensionality present. Of the 52.2%, 3.7% would again fall into correct patterns by chance. The total percentage of individuals that we would expect to find fitting the exact scale patterns would be 51.5%. If this theoretical percentage is compared with the obtained 25%, it becomes clear that, if the assumed

reliability adequately represents the data, more than one dimension is being tapped.

It is suggested here that the expected occurrence of *scale* patterns be calculated assuming (a) independence among questions and (b) uni-dimensionality with a certain degree of unreliability. The obtained percentage of individuals who exactly fit the scale patterns could then be compared with these theoretical percentages using chi square tests. Estimates of the actual reliabilities of the questions used would, of course, make the interpretation unequivocal.

It is the belief of the author that except for the most simple variables, uni-dimensionality will not be found to exist in connection with the measuring instruments which social scientists can construct at present. It would appear futile to insist upon uni-dimensional scales or to make very much of distinctions between scales which possess different "degrees of uni-dimensionality," such as the distinction between scales and quasi-scales.

Scale analysis still provides the investigator with a good technique for scale construction and a means for determining quantitatively the extent to which his data depart from the ideal of uni-dimensionality. Such knowledge should help the investigator considerably in interpreting his data.

GENERAL EVALUATION OF "SCALE ANALYSIS"

Thus far the technique of "scale analysis" has been used mostly with the armed forces. Guttman (5), and McNemar (7) both recommend its use in attitude testing and in public opinion research. Limited experience with its use in public opinion research with civilian populations has tended to show that it becomes an unwieldy instrument. In interviewing people, maintaining good rapport and the interest of the interviewee are important considerations. The process of asking as many as ten questions, all of which are, to a large extent, rephrasings of the same thing, is a considerable strain on relations between interviewer and interviewee. Most of those engaged in this type of research will probably find the inclusion of a series of questions which could be subjected to scale analysis not feasible from practical considerations. In connection with standard paper and pencil attitude tests, however, scale analysis offers the promise of considerable improvement in our measuring instruments.

Many claims have been made by the proponents of "scale analysis." These claims are based upon the mathematics of uni-dimensional measuring instrument. Examples of these claims are (4):

Scale scores provide an invariant quantification of the attributes for predicting any outside variable whatsoever.

In imperfect scales, scale analysis picks out deviants or non-scale types for case studies.

We are assured that if a person ranks higher than another person in a sample of items he will rank higher in the universe of items.

These statements may tend to give spurious impressions of the power of the measuring instrument and may create in the user of scale analysis a greater confidence in his instrument than is justified by the conditions of measurement. Non-scale types may occur because actually more than one dimension is being measured or may occur simply because of unreliability of the measuring instrument. Even if a perfect scale were achieved these claims would all be limited by the degree of reliability of the measuring instrument, that is, of the questions asked.

It might be convenient simply to remember that a uni-dimensional scale of very high reliability would have about the same properties as, for example, the common 12-inch ruler. These properties are, of course, quite important, but there is no need to exaggerate them.

If as is suspected uni-dimensional scales do not at present occur even to a good approximation, then these various claims can be mostly ignored as far as the use of the scale is concerned.

It should be emphasized again, however, that the technique of scale construction and evaluation as represented by "scale analysis" seems to be an excellent technique for use with paper and pencil tests or other instances of measurement where the situation permits the inclusion of several questions centering about the same topic. Where used it will provide useful information about the degree of departure from uni-dimensionality.

BIBLIOGRAPHY

1. CONRAD, H. S. Some principles of attitude measurement: A reply to "Opinion-attitude methodology." *Psychol. Bull.*, 1946, **43**, 570-589.
2. GOODENOUGH, W. H. A technique for scale analysis. *Educ. Psychol. Msmt.* 1944, **4**, 179-190.
3. GUTTMAN, L. The quantification of a class of attributes: A theory and method of scale construction. In P. Horst *et al.*, *The prediction of personal adjustment*, Soc. Sci. Res. Council Bull. No. 48, 1941, 319-348.
4. GUTTMAN, L. A basis for scaling qualitative data. *Amer. sociol. Rev.*, 1944, **9**, 139-150.
5. GUTTMAN, L. Scale and intensity analysis for attitude, opinion and achievement. Mimeographed, 1945.
6. GUTTMAN, L. The Cornell technique for scale and intensity analysis. Mimeographed, 1946.
7. McNEMAR, Q. Opinion-attitude methodology. *Psychol. Bull.*, 1946, **43**, 289-374.
8. NOLAND, E. W. Factors associated with absenteeism in a south-central New York state industry. Doctoral

- Dissertation, Cornell University, 1944.
9. NOLAND, E. W. Worker attitude and industrial absenteeism: a statistical appraisal. *Amer. sociol. Rev.*, 1945, 10, 503-510.
 10. *The desire of enlisted men for post-war full-time schooling: an example of a scale.* Research Branch, Information and Education Division, Army Service Forces. Mimeographed, 1943.
 11. *Experiments on the measurement of the intensity function and zero point in attitude analysis.* Research Branch, Information and Education Division, Army Service Forces. Report D-1, Mimeographed, 1945.
 12. *Questions and answers about scale analysis.* Research Branch, Information and Education Division, Army Service Forces. Report D-2, 1945.

MANUAL OF CHILD PSYCHOLOGY

A SPECIAL REVIEW*

ROGER G. BARKER

Clark University

The appearance of the *Manual of Child Psychology* was a major event in psychological publication in 1946. The announcement of a successor to the *Handbook of Child Psychology*, published in 1933 and long out of print, was greeted with enthusiasm, and with the hope that it would prove as useful and influential as its predecessor had been. Many who studied the *Handbook* as students opened the *Manual* with eagerness, for here was a chance to view as a whole the changes in research on child behavior of these 13 years so momentous in science and society.

Some differences between the *Manual* and the *Handbook* are readily apparent as the following comparisons show:

	<i>Manual</i>	<i>Handbook</i>
Approximate length of text	550,000 words	400,000 words
Approximate length of bibliographies	4,400 references	2,800 references
Number of chapters	19	24
New chapters added to the <i>Manual</i> :	Chapters of the <i>Handbook</i> omitted from the <i>Manual</i> :	
<i>Animal Infancy</i> (Cruikshank)	<i>Locomotor and Visual Manual Functions in the First Two Years</i> (Shirley)	
<i>Physical Growth</i> (Thompson)	<i>The Social Behavior of Children</i> (Buehler)	
<i>Environmental Influences on Mental Development</i> (H. E. Jones)	<i>Children's Philosophies</i> (Piaget)	
<i>The Ontogenesis of Behavior</i> (Gesell)	<i>Speech Pathology</i> (Travis)	
and <i>Maturation of Behavior</i> (McGraw) replace Gesell's former chapter on <i>Maturation and Patterning of Behavior</i>	<i>Eidetic Imagery</i> (Kluever)	
	<i>The Physiological Appetites</i> (Blatz)	
	<i>The Child of Special Gifts and Special Deficiencies</i> (Hollingworth)	
	<i>The Child with Difficulties of Adjustment</i> (Blanchard)	
	<i>Birth Order</i> (H. E. Jones) (Incorporated into the new chapter by Jones)	

The quantity of child psychology has increased in these 13 years: in the *Manual* roughly 20 percent fewer topics require 35 percent more

* CARMICHAEL, LEONARD. (ED.) *Manual of child psychology*. New York: John Wiley & Sons, 1946. Pp. viii+1068.

space and refer to over 50 percent more publications. The quality of child psychology has increased too. The *Manual* reports greatly improved techniques and scientific standards, and careful checking and rechecking of 1933 results. In comparison with the *Handbook*, which appears adolescent in retrospect, the *Manual* is a mature, responsible production in the fields it touches.

But a strong impression made by the *Manual* is one of sameness and familiarity. There can be no doubt about it; Child Psychology, 1946 (according to the *Manual*) is surprisingly like Child Psychology, 1933 (according to the *Handbook*). It is true that the old material is more soundly grounded than when reported in the *Handbook*, and perhaps this is all one should expect. Yet these 13 years seemed to be momentous as they were happening. We thought new discoveries were being made, and it seemed that new methods, new ideas, and new attitudes and practices were being developed. One can picture how a similar book in genetics, biochemistry or atomic physics would compare with a 1933 predecessor! We thought Child Psychology, too, was a lusty, growing, rebellious youngster. So we ask, is this report true?

Let us consider how some of the growing points of child psychology are treated in the *Manual*. Take play, for example. Play has become a central technique of clinical child diagnosis and therapy, it is an accepted experimental method, and it has been the object of fruitful investigation. The index of the *Manual* cites play 14 times, but reference to the text reveals that all of these considerations are extremely brief. When we look under the names of those who have contributed to achievement in this area (for example, Bender, Erikson, Lerner, Levy, MacFarland, Murphy, Sliosberg, Stone) we find a similar summary treatment. Here is one of the active areas of child psychology which the *Manual* treats almost incidentally.

We find the same kind of treatment of social behavior. Social behavior has been another productive area: important techniques and concepts have been introduced, and valuable studies made (for example, H. H. Anderson, Criswell, Doll, Horowitz, Isaacs, Jennings, Lippitt, Moreno, Parten, Partridge, Redl, Slavson, Tryon, M. E. Wright, Zeleny). But a student would certainly gain the impression from the *Manual* that social behavior is a moribund aspect of modern child psychology.

Child personality also receives scant treatment. Although Lewin's work is well represented, the contributions of Blos, Davis, Frenkel-Brunswik, Anna Freud, H. E. Jones, MacFarlane, Murphy, Sanford, Symonds, Zachry, for example, are virtually ignored. Personality and

behavior aberrations in children are hardly mentioned. Sigmund Freud is named 15 times, while fifty other contributors are mentioned more frequently.

The interrelations between physique and behavior in children (physiological, psychosomatic, and constitutional) have received rewarding attention in recent years, yet they are given only fragmentary consideration in the *Manual*.

Recent methodological advances have been made in connection with multi-discipline and longitudinal studies, projective tests, play, interview procedures, sociometric techniques, experimental studies of complex behavior and the factor analysis of data. Save for experimental studies, however, although all are mentioned, they fail to stand out from the other methods used with more or less success in the last 40 years. Since the science of any period is a product of its current methods, it is especially unfortunate for the young scholars who may use the *Manual* that modern methodology is not given more prominence.

In the case of theory of child behavior, we find ourselves again confronted with lacunae. One gains the impression that child psychology during the last 13 years has not been concerned with theoretical problems, and that there have been few serious differences in viewpoint. Aside from Lewin's chapter, theory is seldom mentioned. The reader would have to be intuitive to appreciate that a series of battles of viewpoints have been raging across the child psychology front from the school room to the laboratory (differentiation vs. integration, whole vs. part, Gestalt vs. association, discipline vs. freedom, nature vs. nurture, pure case vs. statistical average, directed vs. non-directed therapy, psychoanalysis vs. all comers, training vs. gratification, test vs. clinical impression). Clinical contributions to the theory of children's behavior are not included.

We have to conclude that the *Manual* gives a conservative, almost old-fashioned picture of child psychology by omitting or inadequately reporting some of the most important developments. A careful examination of the text shows that some of these developments are mentioned, but that they are not placed in perspective. Currently important problems and procedures do not stand out; issues are not clearly stated; results are not strictly evaluated. The *Manual* gives the impression of a catalogue, rather than a narrative.

However, these imperfections should not be over-emphasized. The excellences of the volume should be equally stressed. To some, the general conservatism of the *Manual* will be considered its chief virtue. No book can serve all functions or satisfy all tastes. In fact the effort

to do this appears to be one source of the *Manual's* deficiencies. The 1933 *Handbook* could barely accommodate in an omnibus volume the bulk of the material then available and retain some of its variety and individuality; but Child Psychology, 1946, is too extensive to be packed between the covers of this kind of a book without such fractionation and compression that its significance is greatly reduced and its character largely destroyed.

The *Manual* performs in some degree the functions of a number of specialized publications, the functions, namely, of a *reference work*, a *critical research survey*, a *yearbook*, a *bibliography and series of abstracts*, and an *advanced text*. According to the preface, the *Manual* is "an advanced-level textbook," but internal evidence indicates that a number of the contributors had these other kinds of publications in mind to various degrees as well, publications with characteristics that seriously conflict when the attempt is made to include them all in a single volume. Thus, the vast detail which is a virtue in a reference work interferes seriously with the book's usefulness as a text. As a text, the *Manual* requires introductory material which spoils its effectiveness as a critical research survey for mature scholars. The requirements of completeness of coverage and non-evaluation in a bibliography and abstract are at odds with the need for selection, evaluation and organization in a textbook or a research survey. In a book for research scholars, emphasis upon undeveloped but potentially fruitful problems, and unverified but suggestive methods, is a virtue but in a text these are liabilities. The inevitable slowness of producing a book of this size and with this number of contributors seriously interferes with the possibility of reporting current developments.

An additional handicap with which the *Manual* had to contend was wartime restrictions on publication. Because of this its release was apparently greatly delayed; consequently there are not many citations of literature dated later than 1942. For this reason, its immediate revision is desirable. When this is done, it is to be hoped that the *Manual's* special function will be unambiguously defined. We are at the point in psychology where reference works, research surveys, yearbooks, and advanced texts are badly needed, but it is clear from the *Manual* that we are beyond the stage when a single volume can accommodate such diverse materials. These are tasks which some divisions of the American Psychological Association might well undertake for their specialties.

The order of the chapters in the *Manual* and brief comments upon them follow:

Methods of Child Psychology (John E. Anderson) is an expanded and

greatly improved version of the same chapter in the *Handbook*. Especially welcome are discussions of sampling problems, design of experiments, actuarial and individual prediction, genotype and phenotype, and statistical and psychological significance. However, this chapter suffers much from the limitations of this kind of publication. Under the necessity of touching upon everything from simple to complex and from new to old, the chapter becomes a conducted tour where much is viewed but where nothing can be examined.

The Onset and Early Development of Behavior (Leonard Carmichael) is an expansion of the chapter by the same author in the *Handbook* entitled *Origin and Prenatal Growth of Behavior*. The previous article has become a classic and the present one continues on this same high level. However, one wishes for a book on a subject of this scope; in the present article there are an average of 140 words of comment per research cited. This has been a relatively active field in recent years. The chapter is 20 percent longer and has 43 percent more references than the earlier one; 40 percent of the 500 references are dated 1933 or later. In view of these changes, it is disappointing to find that the issues remain almost exactly where they were in 1933. The author comes to the identical conclusion in 1946 regarding the central issue that he did in 1933, namely "in regard to the related processes of individuation and differentiation of behavior it seems that as yet, at any rate, it is better to record as unambiguously as possible the responses that can be made by a fetus at any stage rather than to attempt to fit all developmental change into one formula."

Animal Infancy (Ruth M. Cruikshank) is a valuable addition to the volume. Here is a summary of research in a field that has been tilled little, but which has important potentialities. It is in the finest and most valuable tradition of scholarly research reviews to define and survey such frontier areas.

The Neonate (Karl C. Pratt) is an extensive revision and expansion of the same chapter in the *Handbook*. The present contribution is almost twice as long as the original article; 45 percent of the 350 references are dated 1933 or later. While the 1933 review presented a framework, here is the completed structure. On the descriptive level, it appears that the neonate's behavior has been fairly well mapped. It is to be hoped that the frontiers of research will soon be pushed to matters of causation, and from the more physiological to the more psychological problems. The work of Ribble (who is not mentioned in the *Manual*), for example, opens important problems deserving careful investigation. This is another research survey which at times approaches an annotated bibliography (100 words of comment per reference).

Physical Growth (Helen Thompson) is a new contribution. To the reviewer's mind, this chapter is out of place. First, as the author states, the material is so vast it cannot possibly be covered adequately. Second,

the crucial problem for psychologists in this connection is the nature of the relation between physique and behavior in childhood, and it is to this that a chapter should be directed. Although this relation has not yet been extensively investigated, it would seem that we have arrived at the point where a short chapter after the manner of *Animal Infancy* is possible.

The Ontogenesis of Infant Behavior (Arnold Gesell) is written to quite a different pattern from the other chapters. It is a stimulating essay presenting an interpretation of developmental data in the form of some descriptive generalizations. This essay will be of interest to mature scholars who will argue long and loudly whether these ingenious generalizations are true and whether they are productive of further research and understanding.

Maturation of Behavior (Myrtle McGraw) covers the maturation vs. learning problem. Few questions in psychology have had such a turgid course, and here is another competent effort to clarify the issues. Despite the author's protestations of the falsity of the dichotomy, she finds it necessary to wrestle at length with a systematic statement of the relation between learning and maturation. It may be that this was inevitable when the task of writing a chapter with this title was accepted, for it implies that a general solution is possible, which then has to be laboriously disavowed. Perhaps in the next *Manual* the author's conclusion will be heeded and we can have chapters on the *Effects of Training upon Early Behavior Development* and *Neural and Physiological Correlates of Behavior Development in Children*, and save ourselves the pain of struggling with the monster of interdependency again.

Learning in Children (Norman L. Munn) is almost twice as extensive as the 1933 chapter by Joseph Peterson. Of the 340 references, 36 percent are dated 1933 or later. In the face of great difficulty, due to the chaotic state of learning studies, the scope of the problem, and the lack of studies directly comparing the learning of children and adults, the author has reviewed a great number of studies of conditioning, acquisition of motor skills, memorizing and problem solving which use children as subjects. He concludes that beyond the very early childhood years when the handicaps of neuromusculature immaturity have been largely overcome, practically all differences in the learning of children and adults may be attributed to differences in motivation and in previous experience.

The Measurement of Mental Growth in Childhood (Florence L. Goodenough) is relatively little changed from the *Handbook* version. Of 109 references 34 percent are dated 1933 or later. Additions include the work of Kelley, Thurstone, Thompson, and McNemar on the factor analysis of intelligence test scores, and Bayley on mental growth. This impresses one as being written in the best tradition of an advanced text.

Language Development in Children (Dorothea McCarthy) is at the

same time one of the most active and hopeful, yet disappointing, aspects of child psychology reported in the *Manual*. The present chapter is over three times the length of the *Handbook* chapter. Of the 480 references, 40 percent are dated 1933 or later. Here is an active field that touches the basic concerns of child psychology: emotion, personality, intelligence, social development, behavior aberrations, education, neurology. However, few of these larger aspects of language have been subjected to investigation. The problems have been defined within a narrow, technical frame, so that language has been largely divorced from its significance as behavior.

Environmental Influences on Mental Development (Harold E. Jones) is a new contribution to the *Manual*. It is packed with information. Here is an adequate reflection of a preoccupation of the times; sixty-two percent of the 235 references are dated 1933 or later. Of all the chapters in the *Manual* this one probably suffers most from the difficulties imposed by the volume. Under the necessity of covering material ranging from I.Q. constancy to diagnosis of identity in twins, and of writing on all levels from pointing out the possibility that coaching affects test scores to discussing the phenomenon of statistical regression to the mean, it has been difficult to present a sharp picture of the issues involved. The chapter will be of great value as a review and commentary for students who bring considerable technical background information, and to those who are able to continue their study in the literature cited.

The Adolescent (Wayne Dennis) meets the issue of the *Manual's* ambiguity of function by sharply limiting the scope of material covered. The psychology of adolescence is restricted to the effects of biological adolescence upon behavior. This makes it possible in the space available to give an adequate review of the research upon this problem.

Research on Primitive Children (Margaret Mead) is a paper for research scientists. It is a stimulating discussion of the methodology of personality studies in primitive cultures, and one that is almost equally applicable to field studies of children within our own society. This chapter provides a sample of the value of a volume prepared for scholars with no concessions to immature students. A unique and valuable departure is the inclusion in the bibliography of citations of research in progress.

Character Development in Children—An Objective Approach (Vernon Jones) is a revision of the chapter by the same author in the *Handbook* entitled *Children's Morals*. This has been a relatively active field; 55 percent of the 140 references are dated 1933 or later. Perhaps because of the nature of the material, this chapter combines better than most the textbook and research-survey functions.

Emotional Development (Arthur T. Jersild) is a systematic review of objective studies of emotion in children; 51 percent of the 238 citations are 1933 or later. It appears that the frontiers of this most wild and

difficult region of child psychology are slowly being pushed back. The picture of healthy progress here given is probably an underestimate by reason of two limitations under which the author has worked: 1) the exclusion of all studies not reaching completely to acceptable standards of scientific method and 2) the separation of emotion from other aspects of the behavior of children. It is obvious that some of the potentially most fruitful developments of recent years are as yet outside the bounds of acceptable science and have had to be omitted. The fractionation of the child and the separate consideration of intellect, language, character, learning, emotion, etc. probably is most limiting and misleading in the case of emotion.

Behavior and Development as a Function of the Total Situation (Kurt Lewin) is the only chapter in the *Manual* written within a definite conceptual frame. However, it is by no means a speculative discussion. Almost all of the 156 references refer to experimental work; 79 percent of the references are dated 1933 or later. This is the most complete statement of Lewin's position available at the present time. Like so much of Lewin's writing, this paper suffers at points from extreme conciseness.

The Feeble-Minded Child (Edgar A. Doll) extends the treatment of the same subject by Pintner in the 1933 *Handbook*. The material has been organized to emphasize standpoints rather than to survey the detailed literature. The topics covered are *definition, classification, incidence, characteristics* and *causation*. Forty-six percent of the 164 references are dated 1933 or later.

Gifted Children (Catherine Cox Miles) is in the pattern of the chapter of the same title by Terman and Burks in the *Handbook*. The framework there erected has been filled in with remarkably little alteration. Valuable new material on the later history of Terman's subjects is reviewed. There are 330 references; 42 percent of them are dated 1933 or later.

Psychological Sex Differences (Lewis M. Terman and Associates) is a thorough, critical review of the widely dispersed data bearing upon this topic. It will be an important starting point for a long time to come for research scholars concerned with sex differences.

The *Manual of Child Psychology* impresses one with the number of raw facts that have been reliably observed and recorded and with the number of interrelations that have been described. This very wealth of data is becoming a burden. In the absence of a strong framework of theory to organize and subsume such an array of facts, and to guide further observation and experimentation, we are in danger of being engulfed. As a theoretical structure is developed we can expect the definition of child psychology to change from one based on chronology (that part of general psychology which uses chronologically young individuals

as subjects) to one based upon psychological criteria. When this occurs, child psychology will become less disorganized and less encyclopedic and the task undertaken by the *Manual* will be less formidable. This is the basic difficulty confronting the *Manual*: lack of concepts and hypotheses appropriate to the empirical content of child psychology. As a compendium of a great amount of this content, and as a publication base in terms of which more specialized books may be produced, the *Manual* will have a great and permanent value.

RESPONSE TO CRESPI'S REJOINDER AND CONRAD'S REPLY TO APPRAISAL OF OPINION-ATTITUDE METHODOLOGY

QUINN McNEMAR

Stanford University

The rejoinder of Crespi (4) and the reply of Conrad (2) to our appraisal of opinion-attitude methodology (9) necessitate a published response to a few points, particularly those having to do with facts. We will first consider three of the points raised by Crespi.

1. We acknowledge a serious error in reporting Cantril's material on the reliability of the polling procedure. Though we correctly reported the only figure for the reliability of opinion questions, we did confuse the data regarding factual-type questions.

2. It is said (4, p. 563) that we "overlooked" at least three articles on reliability, and that if we "had seen" these we "would have appreciated the injustice" of our remark that in opinion research individual "reliability has been *practically* ignored by the users of the *single* question technique" (9, p. 314; our present italics). We did, indeed, fail to report on the three cited papers, but not because they were unseen. We concur with Crespi when he says, "Interested readers are invited to check the facts here for themselves" (4, p. 563). What are those facts, and are they pertinent? Jenkins' paper (7) is concerned solely with questions of the type "What brand did you last purchase?," not with opinion questions. Section D of Dodd's report (5) contains empirical findings on reliability—of questions pertaining to radio listening habits—with no specification as between factual and opinion questions. Now it happens that we had also seen the 148 page, security-marked "confidential," document referred to by Dodd (5, p. 266). This does contain information on the reliability of opinion type questions, but unfortunately the data are not analyzed and presented in a manner which makes it possible to ascertain *how* reliable the questions were. King's article (8) deals with 15 questions which are "idea-centered," hence atypical as regards ordinary single question procedures. Furthermore, detail is lacking for evaluating his report that the agreement between results secured from 25 cases by two interviewers, with a two week lapse between interviews, ranged from 60% to 100%.

3. Crespi says (4, p. 567) that our handling of validity is "plain contradiction," which results in "Three errors—an error of fact, an error of logic, and an error of unjustified condemnation." In proving all this, he quotes from p. 315 of our appraisal. At the risk of tiring the reader, we present the quotation along with parts (here bracketed) omitted by Crespi:

One might have expected that Cantril's volume (1) would have tackled the problem of opinion question validity but such is not the case. [Evidently "the serious problems encountered in every phase of the polling operation" (1, p. viii) do not include what the writer considers to be foremost in importance, namely, validity and reliability for the individual response.] The volume does touch on some of the factors which contribute to [unreliability and] invalidity [but a frontal attack is sorely needed].

In what way is any of the above "an error of fact"? We urge the reader to search Cantril's volume for a validity study. Then having found none, he will see that the "contradiction" and the "error of logic" consist merely in having reported, in a credit line, that some of Cantril's material has a bearing on the problem of validity. Crespi correctly gives the *location* of this material when he states that "the entire point of chapter 1 and, in part, chapter 2 of Cantril's volume is the validity of questions" (4, p. 567). Actually, chapter 1 on the "Meaning of Questions" contains no mention of validity, and in chapter 2 on the "Word-ing of Questions" validity is mentioned only in the discussion of the "absence of any objective criteria of validity" (1, p. 23).

We now turn to five specific points in Conrad's reply and then we shall discuss briefly a method used in his evaluation.

1. Regarding the question of item selection, i.e., the *aposteriori* singling out of particular items for study, we too are "at a loss to understand" (2, p. 581) how we overlooked an entire paragraph which not only definitely negates our statement that no reason was given for choosing items, but also supplies information in line with our surmise about possible capitalization on chance. We hereby tender our apology to Drs. Sanford and Conrad for this oversight.

2. It is evident that Conrad does not (and need not) agree with us on issues having to do with scales, uni-dimensionality, and reliability, but when he defends his scales by claiming that the average inter-item correlations of .08, .12 and .07 for his war-optimism scales are "as high as [those] among items of acceptable intelligence tests" (2, p. 572, with repeats of the same idea on pp. 575 and 586), we wonder what empirical data on acceptable intelligence tests he had in mind. He states, without giving a source, that "intelligence-test items ordinarily correlate with the *total* test only about .30-.50" (2, p. 572). Now to some figures. The correlations for items vs. total score for the 1937 Stanford-Binet Revision (10, pp. 176-185) range from .27 to .91, with 189 of 243 coefficients at or above Conrad's upper value of .50. The mean of these 243 correlations is .61. In the case of the Terman-McNemar Test of Mental Ability the average item-total score correlation is .53 for r 's based on 1200 subjects (13, p. 2). As to inter-item correlations, the average value for the 1937 Stanford-Binet Revision (10, p. 156) is about .38, and from the above mean of .53 it can be inferred by use of Richardson's Formula 7

(11, p. 70) that the average intercorrelation among the 324 items of the Terman-McNemar test is near .28.

One other point on uni-dimensionality: we agree with Conrad's argument (2, pp. 572-573) that for prediction purposes it is permissible to combine in an index anything that improves prediction, but we were not discussing the problem of prediction. We had questioned the meaning of total scores based upon items of near zero intercorrelation and had criticized as meaningless an index of "life satisfaction" based on four uncorrelated components. Since no outside criteria were available, the problem in both these cases was that of scaling not prediction; hence Conrad has confused the issue by bringing in an irrelevant argument. The reader may refer to Guttman (6, p. 148) for a clear-cut statement of the necessity of keeping in mind the distinction between these two closely related topics.

3. Conrad thinks we are "seriously in error" (2, p. 577) in stating that when an item has a reliability of .25 it can be said that 75% of the obtained response variance is due to measurement errors. Now if we understand Conrad's argument against this well-known statistical fact, he is defining an individual's *current* attitude in terms of the obtained response as though the response were not subject to error. He says that "Current attitudes, thus measured, can be denied validity only if it be assumed that item-responses are indeed evanescent" (2, p. 577). One need not assume evanescence if Conrad's reliability estimate of .25 is correct—such low reliability conclusively proves the assumption.

4. Regarding part-whole correlation, Conrad says that "McNemar is mistaken" (2, p. 578). This is proven by a mathematical derivation which begins with the supposition that "if the set of 10 items had a higher standard deviation than the set of 14." This supposition is contrary to the facts: the 10-item scale had an S.D. of 4.1 and the 14-item scale had an S.D. of 5.9 (3, p. 293). Consequently, if our original contention needed an algebraic proof it has now been supplied by Conrad.

5. Conrad (2, pp. 581-583) counters our criticism of his failure to apply statistical tests by saying that good hypotheses can easily be buried by the mechanical application of significance tests. We agree, but we fail to see how that argument justifies ignoring entirely such tests, especially when doing so leads to the conclusion, concerning "one of the most discriminating questions," that a relationship exists which (when computed by us) turns out to be a mere correlation of .06 for 88 cases. In further defense, Conrad states that "The reader will recognize the 'clinical,' 'hypothesis-hunting' nature of the remarks" which describe briefly the procedure which he and Sanford used. This "hypothesis-hunting" consists of beginning with the hypothesis that two variables are correlated, collecting data, making a scatter diagram (from which it should be immediately obvious that the hypothesis is not borne out), next examining the extremes and finding that the median Y-value for

the high and low X-groups "are the same," then looking at the extremes again and noting that the frequencies in the corners of the scattergram are actually 3 and 0 for the high as opposed to 1 and 2 for the low group, and finally basing the conclusion on 6 of the original 88 cases. Any reader who thinks the foregoing an exaggeration is referred to the printed report (12, pp. 13-14).

6. Aside from the above five factual-type points, we would like to mention a few factors that tend to make Conrad's reply sound like a convincing rebuttal to our appraisal. Repeated reading of his discussion along with checking back to our original paper led to some interesting observations. We noted the frequent occurrence of the debater's "my opponent himself has said," along with the stratagem of giving partial quotations and/or ignoring the context from which quotations were lifted. It is not for us to say whether the use of such forensic devices should be discouraged in scientific writing, but in the present controversy it is not only our right but also our duty to point out specific instances which are apt to mislead the reader who does not have time to do detective work.

a. In connection with the problem of the comparison of dispersions on single items, Conrad's discussion (2, pp. 576-578) ends with "We are inclined, rather, to agree with McNemar's statement at another point in his review, that—," and then he gives the following quotation *without* the part which we here bracket: "[Another advantage of scaled opinions is the fact that the] variation within groups indicates the relative homogeneity of groups in their opinion about an issue" (9, pp. 327-328). The reader, who is led to believe that we had been inconsistent, would never guess that the quotation had been removed from our context of scales to bolster an argument in a context of single items.

b. Another out of context quotation, from which it would appear that at one place we sanctioned that which we disallowed at another, may be found on p. 575 of the reply. Our criticism of the assigning of absoluteness to the so-called neutral point of a scale is countered with "As McNemar himself has said," followed by a quotation from a section in which we were arguing against, not for, *such* usage of scales.

c. At the top of p. 575 a quotation purporting to show our rejection of a certain dictum concerning requisite reliability is followed by another quotation, which is characterized as unimpressive rhetoric and said to be the only reason which we offered for our view. In this case the first quotation stops short of a part so relevant that had the sentence been given in full the reader would have grasped our real reason for questioning the dictum.

d. On pp. 573-574 a quotation and subsequent discussion would lead one to think that we recognized "temporarily" a situation for which our tolerance 61 pages later "seems to have vanished," but the quotation

did not endorse the thing (summing scores on components) being discussed at the later place.

e. On p. 579, "McNemar's own criterion" for a scale is cited as supporting an argument; reference to the original (9, p. 294) indicates that we were not speaking of criteria, but of what attitude measurement permits.

f. The *failure* to consider statistical significance in checking hypotheses (see point 5 above) is supposedly supported in part by a "McNemar himself has said" quotation having to do with the dangers involved in using small samples. We never suggested that these dangers in significance tests could be surmounted by the simple expedient of ignoring the tests.

g. On p. 580, "McNemar has himself objected to" is followed by a quotation taken from a paragraph in which we thought we were merely listing some conclusions, not objections.

h. On p. 585, Conrad says "The following quotation illustrates the unsympathetic 'slanting' of McNemar's account." Then a passage from one of several closely related paragraphs (9, pp. 334-336) is given. By starting and stopping at the right places, the illustration possesses a glimmer of plausibility, and from the subsequent discussion of the passage the reader is supposed to see that "McNemar is wrong when he implies that Cantril has erred." Before accepting either of Conrad's verdicts, the reader will do well to turn to the context to see what was not quoted.

We have not herein dealt with all the things which seem to annoy Crespi, nor have we bothered with many points on which we disagree with Crespi and Conrad. The preparation of this response would have been greatly facilitated if it had been necessary only that we admit our errors.

BIBLIOGRAPHY

1. CANTRIL, H. *Gauging public opinion*. Princeton: Princeton Univ. Press, 1944.
2. CONRAD, H. S. Some principles of attitude-measurement: a reply to "Opinion-attitude methodology." *Psychol. Bull.*, 1946, 43, 570-589.
3. CONRAD, H. S., & SANFORD, R. N. Scales for the measurement of war-optimism: I. Military optimism; II. Optimism on consequences of the war. *J. Psychol.*, 1943, 16, 285-311.
4. CRESPI, L. P. "Opinion-attitude methodology" and the polls—a rejoinder. *Psychol. Bull.*, 1946, 43, 562-569.
5. DODD, S. C. On reliability in polling, a sociometric study of errors of polling in war zones. *Sociometry*, 1944, 7, 265-282.
6. GUTTMAN, L. A basis for scaling qualitative data. *Amer. sociol. Rev.*, 1944, 9, 139-150.
7. JENKINS, J. G. Dependability of psychological brand barometers. I. The problem of reliability. *J. appl.*

- Psychol.*, 1938, 22, 1-7.
8. KING, M. B., JR. Reliability of the idea-centered question in interview schedules. *Amer. sociol. Rev.*, 1944, 9, 57-64.
 9. McNEMAR, Q. Opinion-attitude methodology. *Psychol. Bull.*, 1946, 43, 289-374.
 10. McNEMAR, Q. *The revision of the Stanford-Binet Scale*. Boston: Houghton Mifflin, 1942.
 11. RICHARDSON, M. W. Notes on the rationale of item analysis. *Psychometrika*, 1936, 1, No. 1, 69-76.
 12. SANFORD, R. N., & CONRAD, H. S. Some personality correlates of morale. *J. abnorm. soc. Psychol.*, 1943, 38, 3-20.
 13. Terman, L. M., & McNEMAR, Q. The Terman-McNemar test of mental ability. *Manual of directions*. Yonkers: World Book Co., 1941.

BOOK REVIEWS

SYMONDS, P. V. *The dynamics of human adjustment*. New York: Appleton-Century, 1946. Pp. xiv+666.

Symonds regards this book as carrying on the Columbia tradition of dynamic psychology initiated by Woodworth and Thorndike. He readily admits, however, that his contribution is "not exactly along the lines of its predecessors." Since his text is frankly and enthusiastically psychoanalytic in orientation, many experimental psychologists will regard this as an understatement.

Dynamic psychology is depicted as a study of the whole individual and of how he adjusts to the situations he confronts—how he derives satisfaction of his inner drives from his environment. Process and cause rather than personality status are regarded as the subject matter of the field. In fact, the individual as a total personality, his over-all development and inner organization fall into the background in this volume as processes are developed. This is shown in the treatment of the ego concept which is not discussed in one section but presented fragmentarily throughout the book as various other Freudian concepts are handled nomothetically.

"... Reason and intellect are dethroned as the principal factors in adjustment" the preface states, and "adjustment [is] primarily a matter of the reactions to frustration and the individual's attempt to avoid anxiety. ... The use of reason in meeting problems of adjustment is reached, if at all, only in maturity as the result of high intellectual endowment and the capacity to profit by experience." "... adjustment is carried on through the impulses, emotions, and by means of the mechanisms which are described herein." It would be interesting to know the extent to which members of the profession have reached this same conclusion and the effect in such cases of this view upon their theory of education, their courses and research in psychology.

Symonds has not been known as a psychoanalytically oriented writer. Many who are acquainted with him as the author of an excellent organization of the personality measurement literature in *Diagnosing Personality and Conduct* will find this volume surprising. The differences in interests and viewpoint shown by the two contributions may reflect a trend in psychology as well as the capacity of an alert, scholarly psychologist materially to change his approach in middle life.

The author purposes to represent the conclusions of contemporary research and discussion on the problem of human adjustment. He states that as he combed the literature he concluded that the psychoanalytic contributions commanded the field. The basic theory of Freud and of ego psychology are presented with the refinements suggested by later students including Horney, Fromm, Reik, Fenichel, Deutsch, Alexan-

der, Klein, and Isaacs—writers whose work is too little known to students of adjustment theory. The volume is not written primarily from a background of system and concepts of traditional psychology with psychoanalytic literature interwoven, but rather the converse. The chapters on drives, frustration and conflict allude to papers by psychologists. Many of the other chapters such as "Love and Self-Love," "Guilt and Self-Punishment" and "Miscellaneous Mechanisms" contain very few contributions from members of the American Psychological Association. On the whole the experimental studies on personality mechanisms seem ineffectively presented.

In his preface Symonds anticipates criticism on the score that statements are made without apparent evidence, which makes the book seem dogmatic. He emphasizes in rebuttal that when one makes an intimate and detailed study of human life through clinical methods, it is done under conditions by which it is not possible to duplicate data. Moreover, he contends, "... when relationships are verified by repetition of scores of investigators, relationships between observed behavior and the dynamic processes which motivate it gain credence." This presentation may seem unnecessarily weak to the many students who come through introductory courses which emphasize objective methods and attitudes, in which experiments are cited for each generalization made. Too, the students from nonscientific curricula who will use this text need a cautious approach to broad clinical generalizations, which this text lacks. It is regrettable that Symonds did not take a more critical attitude toward some of the findings as he presents them, and at least acquaint the reader with the methods of obtaining these data and with well-known attacks on Freudian theory and technique. Furthermore, the questions of relative reliability and validity of data might have been raised. The student is at a loss to distinguish fact from theory and hypothesis. Too seldom is he told that additional studies relevant to a given concept are sorely needed.

Statements such as this will arouse criticism: "... there is also a part of the ego that says, 'Do' and 'Don't' that is also unconscious" (237). And many will want to know the degree of validity of assertions such as: "[early frustration in oral experience produces] mouthy folks—verbose in expression, persons who like to pull on a big cigar or heavy pipe, or chew on the stub of a pencil."

The mixed style and irregular organization within chapters will irritate many of the more perfectionistic readers. The integrated pattern of psychoanalytic insights tends to be lost as Symonds attempts to organize mechanically the literature in terms of single mechanisms.

Symonds deserves great credit for his efforts toward the integration of a body of widely varied literature, for his attention to truly significant, complex and somewhat inscrutable problems, for his wholesome psychological treatment of social problems arising from individual

maladjustment, and for his over-all organization of the major concepts of dynamic psychology. This respect is enhanced when his volume is compared with some other publications in the field. The 63 pages devoted to notes concerning the 883 references are a contribution in their own right. This work should give pause to the personnel worker who dispenses superficial advice and recommends social maneuvers to alleviate deep conflicts.

FRED MCKINNEY.

University of Missouri.

BENEDEK, THERESE. *Insight and personality adjustment: a study of the psychological effects of war.* New York: Ronald Press, 1946. Pp. xii + 307.

As a psychoanalyst of the staff of the Chicago Institute of Psychoanalysis, Dr. Benedek is probably fairly orthodox. As the writer of a book, practical and timely, and geared to the immediate clinical problems of today, Dr. Benedek seems, to the reviewer at least, to differ sharply from most orthodox psychoanalysts. With brief acknowledgment to the theoretical substructure of psychoanalysis, the author addresses herself promptly to that large body of personalities today who reflect the tremendous need for personal, postwar readjustment. The veteran is given careful consideration, not as the *species veteran*, but as he is a husband or a father or a son. What he has in common with other veterans—the war itself—is evaluated psychodynamically as an homogenizing influence. Just as carefully does the author evaluate the war as an experience for the wife and the mother of the veteran, the father and the child, the brother and the sister.

To infer from this that the author is focussed on personalities would be unfortunate, for her concern seems to be primarily with situations: separation, continued separation, reunion, personal readjustment. These situations and the behavior they demand are presented each as a psychodynamic equation.

To the persons involved with it in various ways, the war experience served or failed to serve needs, and an analysis of it in these terms is a healthy exercise for those who today must consider problems of readjustment. Dr. Benedek states her hope that this book will be found useful by social workers, clergymen, teachers, counselors, as well as psychologists. That many such readers will be stimulated by her book I have no doubt, for I liked it. It is clear that Dr. Benedek has thought her way through the present world crisis as it is involved in clinical work. Much more than that, she reveals an insight into present-day problems of personal adjustment which must derive from more than psychodynamic theory and thinking alone. She has seen real people trying desperately to meet today's problems.

Just how this book will be used by counselors, it is difficult to guess;

for it is in a sense a compendium of dynamics, codified according to the interpersonal relationship involved. Whether or not the book will serve as such a reference work, the reading of it as a whole will provide an intelligent psychodynamic analysis of the present world crisis as it troubles the individual today. As a book about individuals this is a book in clinical psychology, and I think it has, from this viewpoint, an unfortunate title.

THOMAS W. RICHARDS.

Fels Research Institute.

HEWITT, L. E., & JENKINS, R. L. *Fundamental patterns of maladjustment; the dynamics of their origin.* Institute for Child Guidance, State of Illinois, 1946. Pp. 110.

From the files of the Michigan Child Guidance Institute 500 case records were selected and certain information in these records entered on a code schedule. "Three hypothetical behavior syndromes were suggested including the following items: (1) assaultive tendencies, temper displays . . . etc.; (2) gang activities, . . . stealing, . . . truancy, etc.; (3) sensitiveness, seclusiveness, shyness, etc." These were formalized as (a) unsocialized aggressive behavior, (b) socialized delinquency behavior, and (c) over-inhibited behavior syndrome.

A very large number of tetrachoric and biserial correlation coefficients were obtained between the items and combinations of items. The mass of the statistical analysis is presented in a fashion which the reviewer found confusing. He was unable to see the meaningful forest because of the crowded numerical trees. It is his impression that most of the traits (items) which intercorrelate highly, for example, .67 between seclusiveness and shyness, are self-evident. No clear line of evidence was apparent indicating that other than obvious relationships were found. The evidence may be there, but if so it is obscured by too many statistical tables.

Sample case histories are given to illustrate the personality development of the three syndromes. Dr. Jenkins adds a chapter giving the psychiatric interpretation of this study, together with its implications in guidance and treatment.

The authors concluded that,

This study has served to provide a statistical verification of certain previously established hypotheses . . . concerning the nature and backgrounds of the aggressively delinquent and the pseudo-social delinquent. The situational correlates of the third type of behavior (overinhibited behavior) have been revealed through empirical analysis. In each of the three behavior-situation pattern relationships there appears to be some evidence that not only is the behavior in question provoked by a particular type of frustration but the general pattern of behavior itself is exemplified by other persons with whom the child is in

close contact. Thus the resulting type of maladjustment would appear to be a rational reaction of the child to his distorted environment in a double fashion.

C. LANDIS.

*New York Psychiatric Institute
and Columbia University.*

THORPE, LOUIS P. *Child psychology and development*. New York: Ronald Press, 1946. Pp. xxvi+781.

This book is intended as a text for courses in education and psychology. Many instructors, however, will find it more useful as a reference source for their own use, and that of their more advanced students. The book is long, repetitious, and at times the writing is not clear. In some instances the selection of research materials, and in many cases their interpretation, is extremely biased. For these reasons it does not meet the needs of the average undergraduate.

Each chapter begins with an historical and theoretical discussion of the topic to be presented. There follows a discussion of findings, including a series of presumed typical investigations and a "summary and implications" in support of Thorpe's own point of view. The book is heavily documented through the use of many long quotations from the educational and psychological literature and numerous footnotes. Questions for study and a list of references are appended to each chapter.

Chapters two, three, and seven, *What the Child Inherits, Mental Abilities: Nature and Nurture*, and *Intelligence and How It Develops*, deal with the operation of heredity and environment in the process of child development. Thorpe obviously has strong environmental leanings and is, of course, entitled to his point of view. However, this reviewer questions the use of such terms as "congenital inheritance" (pp. 71-72) and "experiential maturation" (p. 144). Few psychologists today would deny the mutual interdependence of hereditary and environmental influences from the moment of conception. But it seems unnecessarily confusing to use "inherited" for such environmentally produced effects upon the child while *in utero* as those caused by malnutrition, disease, or injury to the mother, even with "congenital" as a qualifying adjective. Nor does it seem justifiable to nullify the term "maturation" which by definition means that part of the developmental process which is impelled from within, by preceding it with a word descriptive of extrinsic factors. Thorpe draws a clear-cut distinction between physical attributes which he concedes to be inherited, and all mental and personality traits which in his opinion have neither structural nor innate bases.

Although both sides of the nature-nurture controversy are presented, researches supporting the hereditarian view are sometimes out-dated

where more recent and better investigations are available. The usual misconceptions by which the controversy is perpetuated are reiterated in this book: accepting test results at face value, after warning others not to do so, assuming that because certain conditions invalidate the use of tests that the function tested has changed rather than that the measuring instrument has been rendered useless, and finally, in lieu of conclusive evidence, holding the view that intelligence per se is improvable because we would like to think that it is so. It is unfortunate that these chapters take up so large a portion of the first part of the book. Many will read no further and miss some of the more acceptable materials presented later on.

Thorpe rejects *instincts* and *drives* in favor of *basic needs* as the underlying motivational forces of child behavior. The child continually strives to satisfy these needs, and society's role is to see that he succeeds in a socially acceptable manner. Since the author recognizes no innate personality limitations, a suitable environment is all that is required to produce children free from frustration and conflict. The individual parent is warned against "over-acceptance" of the child, but apparently there is no limit to the extent to which society-in-general should plan an environment accurately attuned to the child's needs.

Emotions are "stirred-up states of the organism," resulting from failure of the child to satisfy his needs. The traditional (puritanical) point of view, that strong emotions are undesirable and to be avoided, is upheld by the author. Pleasant emotions receive scant attention.

In keeping with Thorpe's strong environmental bent, the best chapters in the book are those dealing with *Effects of Early Home Conditions*, *The Social Education of the Child*, *Safe-Guarding the Child's Personality*, and *Mental Hygiene*. An extensive presentation of theories, terms, and research findings is followed by many sound suggestions by which parents and teachers can improve the conditions under which children live and learn.

For the mature student, capable of making his own selection from the wealth of material presented, the book will be a welcome addition to his sources of reference. He will find useful, also, the numerous tabulated summaries of research studies.

KATHARINE M. MAURER.

University of Nebraska.

SCHROEDER, E. M. *On measurement of motor skills; an approach through a statistical analysis of archery scores.* New York: King's Crown Press, 1945. Pp. xvi+210.

An interest in measurement and testing in the field of physical education activities was the motive for the research reported in this book. Archery was selected as the subject of study on the ground that it is a highly routinized activity carried out under relatively well stand-

ardized conditions; its score is objectively measured, accurately recorded, and commonly accepted as a valid measure of the performer's skill.

Based on data collected over a period of six years from the performances of 258 beginning, 139 intermediate, and 53 advanced archery students at Wellesley College, and limited in scope to the investigation of the behavior of the score, the study is divided into two parts. In Part One, which deals with the trend and reliability of the score in successive lessons, the purpose is to evaluate score units of stated length as tests of group and individual ability and improvement. Part Two concerns the behavior of the score within a single lesson with the object of determining whether successive small units of score reflect the effects of practice and fatigue, and whether such a lesson is long enough to measure individual skill accurately.

Because of the standardized nature of archery, many of the findings of the study—such as, for example, that the sum of the best two scores estimated individual skill more accurately than did the sum of the last two scores of the series of lessons, that the Range score on twenty-four shots was sufficiently reliable to be a useful test of group ability, that the lesson of standard length was too short to measure accurately the skill of the majority of subjects—will have direct application to testing in that field. Taken as a whole, however, the results obtained illustrate and emphasize a fact which is of prime importance in the interpretation of test scores, namely, that a score made at any stated time is likely to estimate ability with varying degrees of accuracy. For the light it throws on the nature of scores, the study should be helpful in suggesting procedures to improve the accuracy of measurement in other sports activities.

The methods by which the various types of problems undertaken were handled are presented in detail. Careful definition of terminology and the adequate treatment given to the statistical techniques employed add to the effectiveness of the presentation. As an exposition of testing problems that are unquestionably fundamental, the book is an excellent contribution.

JAMES M. LYNCH.

Personnel Research Section, AGO.

EDWARDS, ALLEN E. *Statistical analysis for students in psychology and education*. New York: Rinehart, 1946. Pp. xviii+360.

This book is intended primarily for students of psychology. The conventional chapters on scaling, reliability, and validity are omitted on the grounds that they belong in courses on tests and measurements, while consideration of multiple correlation is postponed to more advanced courses. The author begins with a simple review of basic arithmetical computations, such as fractions, decimals, square roots, posi-

tive or negative numbers, and then goes on to the treatment of measures of central tendency and variability. A notable omission is the graphic treatment of distributions, and this is regrettable since some guidance in the graphic presentation of frequency polygons (or histograms) is an essential part of a student's training in elementary statistics if his training is to help him in subsequent laboratory or field work in psychology.

The treatment of fiducial limits and reliability of means, which so often is either garbled or even erroneously presented in some elementary texts, is treated here accurately but in a rather verbose manner. It is gratifying to see that analysis of variance has been finally divorced from the barn-yard imagery which adheres to this term in agricultural and biological textbooks.

The chapter on χ^2 is primarily devoted to the treatment of qualitative or enumerative data and it would seem better to give it such a title. In general, this tendency to deal with the names of special techniques as chapter headings rather than with the contents of these techniques is unfortunate, for it gives the student the feeling that statistics is made up of disparate tools. It would be much better to organize the content according to topic and let χ^2 and comparison of percentages, for example, fall into the same group. Similarly, analysis of variance could well form part of a chapter or a series of chapters on comparison of means and thus give the student an integrated point of view regarding critical ratios, t tests, F ratios, and correlation ratios. This is done to some extent on p. 199, but because of the segregation of the techniques into separate chapters it may appear to be an accident rather than a necessary result. By bringing these methods together, the consistency of different treatments can be demonstrated and the unity of statistical method stressed.

The final chapters on sampling prediction and design of experiments fulfill a basic need in an elementary text.

In general, the text's chief advantage lies in the fact that it draws its material entirely from psychological sources and thus makes the student of psychology feel quite at home within its pages. In some respects, it descends to quite an elementary level, presuming very little, perhaps too little, on the part of the student. In other respects, especially in the chapter on analysis of variance, it goes far beyond the proper limits of an elementary text.

In common with the majority of elementary texts, this book devotes little space to the consideration of even simple curvilinear relationships with the result that the student gets the impression that psychological data are primarily linearly related, and that functional or curvilinear relationships do not exist, are due to artifacts or are beyond the pale of psychological investigation. As a consequence, the student of psychol-

ogy is completely at a loss when he begins to deal with such nonlinear relationships as the Weber-Fechner Law, the many functional relationships in the field of visual and auditory perception, and the growth curves in developmental psychology. The treatment of functionally related variables is no more complicated than that of analysis of variance or the correlation coefficient. Since it is so useful to the student of psychology it should find a place in the elementary texts.

JOSEPH ZUBIN.

*Columbia University and
New York Psychiatric Institute.*

GOTTSCHALK, L., KLUCKHOHN, C., & ANGELL, W. *The use of personal documents in history, anthropology and sociology*. Social Science Research Council Bull. 53. New York: Social Science Research Council, 1945. Pp. xiv+243.

This Bulletin comprises three separate and distinct monographs. Gottschalk gives brief and concise descriptions of historiography and historical method in the use of documents. In similar fashion he deals with the principles of external and internal criticism of documents. These principles are usually more important to historians than to psychologists and social scientists, since the latter seldom have reason to question the authenticity of documents which come to them. Nevertheless, social scientists frequently use historical method, and Gottschalk's suggestions regarding the use of documents should help them to avoid pitfalls in their excursions into historical fields.

Kluckhohn points out that anthropologists, unlike their colleagues in other fields, seldom work with materials which have been produced spontaneously for confidential use. The documents employed by anthropologists are usually written at the instigation of field workers, or are merely records of interviews. Kluckhohn's critical review of such documents, including a number of biographies, is an excellent guide to the literature. The suggestions which he makes for future research, and for making available more of the original and detailed notes from field work, represent no mean contribution to methodology and technique in anthropology.

Angell catalogues and discusses the various ways in which personal documents are used by sociologists, emphasizing that such materials both yield hypotheses and contribute to the verification of hypotheses. He reviews the principal sociological studies of documents under two headings: (1) explanation of historical sequences, (2) contributions to sociological theory. A brief chapter on sociological method is followed by one on suggestions for obtaining, analyzing and interpreting documents and demonstrating the validity of hypotheses from documents.

No brief review can do justice to the importance of these monographs or to their usefulness as guides to users of documents. Nor can they be evaluated properly without considering them in the light of publications of a similar nature by Dollard, Blumer, and G. W. Allport, the latter two of which are authors of previous SSRC Bulletins. The definitions of *personal* document (or *human* document, as the case may be) given by these six writers are so different and distinct that they suggest the analogy of the blind men describing the elephant.

Blumer defines the *human* document as "an account of individual experience which reveals the individual's actions as a human agent and as a participant in social life."

Allport defines the *personal* document as "any self-revealing record that intentionally or unintentionally yields information regarding the structure, dynamics, and functioning of the author's mental life."

Angell, for his purposes, defines a *personal* document as "one which reveals a participant's view of experiences in which he has been involved" (p. 177).

Kluckhohn agrees with Dollard that the subject of a personality sketch or biography "must be viewed as a specimen in a cultural series," but adds that "the analysis will be incomplete unless the universal, communal role and the idiosyncratic components of personality are distinguished" (p. 138).

Gottschalk, in contrasting Blumer's definition of *human* document with Allport's definition of *personal* document, regards these adjectives as tautological, since "every document, no matter how thoroughly the author strove to be objective, must exhibit to a greater or less extent the author's philosophies and emphases, likes and dislikes, and hence betrays the author's inner personality" (p. 13).

These definitions are quoted not so much to emphasize that the representatives of the several academic disciplines are not in complete agreement regarding their fundamental concepts, as to point out that in stating their definitions boldly and precisely these writers have provided the key to an understanding of the distinctive methods and techniques characteristic of sociology, psychology, anthropology, and history applied to documents. In these days when psychologists, social scientists, and historians are being encouraged to pool their concepts and techniques in grand "cooperative projects," there is strong temptation to urge that all of them should get together and agree on fundamental concepts and methods. While such cooperation is desirable, it must be remembered that data, interpretations, and generalizations are in part functions of the concepts, methods, and techniques employed in the research which produced them. Pooling concepts, methods, and techniques may result in data of doubtful value, interpretations devoid of clarity, and generalizations lacking precision and validity, unless the cooperators have scrupulous regard for the various methodologies.

These monographs are excellent guides to such methodologies, at least in so far as the use of documents is concerned.

ARTHUR F. JENNESS.

Williams College.

LILLIE, RALPH S. *General biology and philosophy of organism*. Chicago: University of Chicago Press, 1945. Pp. 215.

The question—What is life?—constitutes the central theme of Professor Lillie's book. That he is well qualified to discuss this problem will be denied by few; that he has been successful in his attempt to formulate an answer will be denied by many.

It will come as no surprise to those who have read Lillie's many contributions to the *Journal of Philosophy*, *Philosophy of Science*, and other scholarly journals, that he endorses a teleological interpretation of life. Such a vitalistic view is somewhat novel in a day when scientific sophistication appears to be synonymous with an acceptance of physical determinism.

Following a brief review of cellular chemistry, Lillie attempts to show that the basic fact of chemistry is that the orderly and complex distribution of molecules requires energy. He feels that this energy must be applied "—directively within the organism," in order to overcome the natural tendency for molecules to become distributed into random and symmetrical patterns, a condition which is presumed to be due to chance. This natural tendency is expressed in the second law of thermodynamics and is a tendency which must be overcome if there is to occur any cellular differentiation. The required energy, which he refers to as an anti-diffusion factor, is, according to Lillie, essentially teleological. That is, it is applied directively with some purpose.

For him, the organism is a psychophysical system in which the final integration of the physical and psychical factors appears to be psychical rather than physical. The physical is characterized by constancy, by permanence, and by a static nature, whereas the psychical is novelty, is change, is process and occurs only in the present. The psychical cannot be observed externally as can be the physical but is instead felt and self-experienced. This is characteristic of one of the concepts of phenomenology. He expressly disclaims a dualistic view but rather feels that the two fields, the physical and the psychical, are grounded in the same fundamental reality. This corresponds closely to the monistic double-aspect view of psychology.

According to Lillie, the entire psychophysical organism acts as if it were the field of a specifically unifying factor, a factor which has constant properties but whose activity is directive and synthetic in essence. This action is carried out in pursuit of definite purpose and "psychic aim" becomes dominant and determines the special direction of what happens in the physical world. For example, Lillie rejects the theory

that the natural selection of purely fortuitous variations (e.g. mutations) is a sufficient explanation of the origin of the entire range of adaptive characters. He definitely feels it necessary to explain evolution in terms of a directive purpose.

The crucial point in Lillie's theory, and one which he quite clearly recognizes as being of vital importance, is how the psychical can operate on and give direction to a physical system when the psychical factor, as such, cannot exert physical force. He attempts to get around this dilemma by placing the locus of psychical control internal to or behind the elementary physical events (quantum transfers). How this solves the dilemma is not clear. It appears to be a restatement of the problem in another way and on another level. While some may feel satisfied that Lillie has found a locus for the psychical factor (the reviewer does not), the way by which it can exert physical force is still an enigma.

How then are we to evaluate this book? Is it possible that the author has stated his views too soon, that he should have waited for more evidence? If Professor Lillie were at the outset of his career, we might reasonably expect him to remain an agnostic until more evidence should become available. But he is not at the beginning of his career. He is already an emeritus professor. Furthermore, his lifetime in the laboratory has given him much experience that could be obtained in no other way. And is it not worthwhile for younger scientists to learn of his experience?

It is probably impossible to correctly say whether Lillie is right or wrong in his assumption of a teleological interpretation of nature. True, he does not have the weight of numbers in support of his position, but for that matter neither has any other scientist who has dared to break with tradition. Surely the history of science has taught us the folly of mere majority opinion. It must also be taken into consideration that there is more hope in Lillie's view than in that of the physical determinist, but, unfortunately, wishing does not necessarily make a thing so. Perhaps the chief merit of the work is in the possibility that it will stimulate further research on the problem from which an unequivocal solution may be derived.

ROBERT P. FISCHER.

University of Florida.

FALES, WALTER. *Wisdom and responsibility—an essay on the motivation of thought and action.* Princeton: Princeton Univ. Press, 1946. Pp. 166.

In this closely-written, highly epigrammatic, and occasionally obscure series of ten essays, an objective theory of value-experiences is presented. Like Koffka, the author holds that we are earlier aware of the subjective values of objects than of their spatial and temporal attributes. These "subjective values" are *motives*; objective values are

mental categories in terms of which motives are "innately" interpreted and judged. Intentions are prior to thoughts, decisions prior to insights. We are able to organize because we are ourselves previously organized and to create because we are creatures. "Everything which has weight in a man's life or meaning in his thinking derives its structure from his final ends" (p. 161).

Individuality itself is nothing but the ability to regulate the growth of motives in such a way that situations become solvable (p. 26). Every individual develops in a chain of decisions which can be understood as evidencing a system of objective values; but our genuine decisions are determined by final ends which we do not even see although we have a relation to them. Personality or "model individuality" is the response to *calls* rather than to *needs*, i.e., we stand for something that is bigger than we are (p. 68).

Educationally, this outlook leads to the important suggestion that it is the function of intuitions to perceive wholes which account for the coherence and meaning of that which otherwise would remain below the threshold of interest and attention. For example, knowledge obtained by way of inference is encyclopaedic and the property of all who are intelligent enough to take possession of it; but more fundamentally, those impulses that lift us up to specific planes of understanding are symptoms of an anxiety which drives us to master the problem "world" in accordance with standards set by our final ends (p. 148). This defines the problem of the educator as the control of the momentum of powers active in shaping the better selves, i.e., the system of demands of other persons—all because human life is essentially obligation.

Superficially, this seems like a contemporary restatement of ethical transcendentalism, but on closer scrutiny it takes on more of the character of a novel form of radical empiricism or field theory that seeks to place the Jamesian marginal in experience on the same or even loftier ontological footing than the focal regions of our phenomenal world. Only in such a context does the assertion that we are least selfish when we are most creative become intelligible. The key suggestion that motivation is ultimately impersonal in the sense that the organism relays forces greater than itself is truly profound, and may compel some useful revisions in the need schemata now popular among applied psychologists and "social engineers."

GEORGE W. HARTMANN.

Teachers College, Columbia University.

SARGENT, W. E. *Teach yourself psychology*. Philadelphia: David McKay, 1946. Pp. 159.

Teach Yourself Psychology is the sort of popular presentation that one hopes will not have a wide audience. The first chapter, consisting of little more than a listing of the names and dates important in the pre-

experimental history of psychology, furnishes some basis for believing that this hope will be justified. In the somewhat more readable discussions that follow, covering among others the topics of instincts, structure of the mind, and dreams, most of traditional experimental psychology is included in a single chapter. This reflects, of course, the author's clinical (and religious) bias. He is still fighting the battle against Watsonian behaviourism, and he fights with the same religious zeal manifested by critics 20 years earlier. His own theory is taken in about equal parts from Freud and McDougall. The tone of the volume is typified by the following sentence: "Further, if man is nothing more than a machine which reacts to the world around him as a typewriter responds to the touch of the typist, how can we explain the visions of prophets and the dreams of seers whose thoughts are unique and in advance of their generation." The concluding sentence contains another example of the same sort: "Its (psychology's) final aim is not merely to state how man thinks, feels, and acts, but how he can do these things much better and more in accordance with the Divine purpose that lies at the back of all things; . . ." If the present volume is typical of the rest of the Teach Yourself series, the publishing effort involved would seem to be quite ill-advised. There is little question, however, that it is easier to make a mistake of this sort in psychology, than, for example, in algebra.

LLOYD G. HUMPHREYS.

University of Washington.

BLUMENFELD, WALTER. *Introducción a la psicología experimental*. Lima: Editorial Cultura Antártica, S. A., 1946. Pp. 417.

For more than a decade Dr. Blumenfeld has furthered the interests of psychology in Peru, and the influence of his prolific and constructive activity has extended over much of South America. His production of this introduction to experimental psychology in Spanish is likely to be of great importance in establishing psychology as an autonomous field in regions where it exists largely in an incidental way in relation to other interests.

The book is a *précis* of experimental psychology, covering in 24 well-arranged chapters all the topics one would expect, supported by the inclusion of very recent data. A full bibliography is drawn from a wide range of experimental literature and convenient tables, clearly printed charts, and excellent diagrams including some plans of laboratory apparatus, are extensively used. In style of language the exposition is consistently clear and economical. Publication was sponsored by the Instituto Psicopedagógico Nacional, which organization, along with the publishing concern, deserves considerable praise for a well-executed undertaking. In a straightforward preface the author tries to make clear to his readers the basic importance of experimental procedures for all psychological work, and thus provides an emphasis that has been

lacking in some parts of Latin America. In discussing such topics as psychoanalysis and personality he shows an enlightened scientific regard for the bearing they have upon experimental psychology. More than half the book is devoted to sensory and perceptual processes, and the psychophysical methods are well presented in their historical contexts as well as in present applications. It is interesting to observe through Dr. Blumenfeld's portrayal the greatly varied utility of rather a small number of scientific principles for building up the total experimental field.

Without introducing any dogmatism of theoretical approach the author quietly and effectively achieves his intention of allowing experimental psychology to speak for itself. The logical background is highly eclectic. And when it becomes necessary to make a break with the somewhat speculative philosophical tradition of Latin-American psychological thinking, the break is always made gently, considerately, and courteously. The strong presentation of a needed emphasis amidst an atmosphere devoid of polemics is a tribute to Dr. Blumenfeld's grounding as a psychologist and skill as a writer.

HOWARD DAVIS SPOERL.

American International College.

LINK, HENRY C., & HOPF, HARRY ARTHUR. *People and books*. New York: Book Manufacturer's Institute, 1946. Pp. 166.

Will the war-time boom in book-buying decrease with the peacetime availability of essential and luxury items? The true guide to the future of books is the reading habits of people, not just the present sales figures. In 1945 the book industry's business reached a half billion dollar total and yet up until that time no one had scientifically investigated the market.

Through the combined efforts of all the major groups of businesses which participate in the production of books, the Psychological Corporation and the Hopf Institute of Management were retained to meet the need for a reliable and accurate index of consumer book reading and buying habits. An extensive survey was organized to answer the main questions about which the publishers and others of the book industry had heretofore been able only to conjecture and theorize.

Four thousand interviews were conducted with consumers between May 21 and June 8, 1945, after a six month period of developing methods and revising questionnaires. Acting as field supervisors, sixty-two psychologists directed the work of the interviewers in 106 cities and towns.

The results of this research show clearly that income is *not* the dominating influence, but rather that formal education is the determining factor in book reading. To be sure, people have price preferences in books, but these are not based entirely on economy; often they reflect

motives which can be appealed to by the makers of both the lower-priced and the higher-priced books.

Book stores, book clubs, and department stores, in that order, are the principal sources for the consumer purchase of books. Purchasers do not mean readers, however. In this study it was found that best-seller lists give no indication of the number of people actually reading the books. During the period preceding this survey the reading of *Forever Amber* equalled if not surpassed the reading of the Bible!

The results show that the population spends twelve times as much time per day on newspapers, magazines, radio and movies as it does on books. The reviewers do not feel, however, as did the investigators, that this necessarily indicates so great an amount of *available* book reading time. We feel that before any such generalization as to potential book markets can be made, the *purposes* for which newspapers and magazines are read and the activities-while-listening of radio users must be ascertained.

Forty-one per cent of the persons surveyed in this study said they own less than 100 books; thirty-four per cent claim to have more than 100 books. The accuracy of respondents estimates was not empirically verified by a sample.

A parallel survey was conducted with dealers, distributors, publishers, and educators. The results corroborated the consumer survey in predicting an expanding market for books.

Two suggestions occur to the reviewers for possible improvement of the study: (1) an analysis of conditions under which the book industry should decide to contract, maintain, or expand future production facilities, so as to utilize more fully the present data, and (2) the development of a predictive formula for future commitments to be based upon the comparison of book sales with survey findings when, as the investigators suggest, surveys of this type are repeated.

NANCY C. COOLEY.

ROBERT H. SEASHORE.

Northwestern University.

CLEETON, G. U., & MASON, C. W. *Executive ability: its discovery and development.* (2nd Ed. Rev.) Yellow Springs, Ohio: The Antioch Press, 1946. Pp. iii+540.

According to the authors the purpose of this book is to report and coordinate the best available information concerning the qualities necessary for proper performance in executive positions. The topics treated range from neuromuscular activity to a definition of democracy. Emphasis is placed on problems of selection and training and the role of executives in labor relations. The treatment appears to be strongest in the last area and less strong when the discussion turns to scientific principles of behavior. Typical of the latter case are the acceptance and use

of terms such as instinct, will, and mental capacity without a clear statement of operations defining them.

The major shortcoming of this volume is that sweeping generalizations, presumably based on research findings, are reported without the necessary information to evaluate the instruments, procedures, statistical tools, and data. A few other criticisms may be listed. The presentation of selection procedures (roughly 125 pages) omits the critical concept of cross validation. Over-emphasis on rating scales leads to the ignoring of more objective measures of proficiency. The authors believe that there is no communality among the traits necessary for success in different executive positions; this position implies a research program for each executive job, certainly an impractical task.

It is the opinion of the reviewer that the attempt to treat this important applied field is admirable and that this book may be useful in presenting elementary notions for the non-psychologist working in this area. It is felt, however, that the book falls short of its stated purpose of summarizing the best available data concerning executive ability.

WILLIAM O. JENKINS.

Indiana University.

BLAIR, GLENN MYERS. *Diagnostic and remedial teaching in secondary schools*. New York: Macmillan, 1946. Pp. xv+422.

Dr. Blair's book has a two-fold purpose: "to supply teachers, principals, supervisors, and superintendents with concrete and practical suggestions for carrying out remedial programs" in secondary schools, and to serve "as a basic text in courses in diagnostic and remedial teaching . . . in teacher-training institutions." Accordingly, the book is divided into three parts. Part I, consisting of the first seven chapters, is devoted to "Diagnostic and Remedial Teaching of Reading." Part 2, consisting of chapters 8 through 11, deals with remedial work in secondary schools, in arithmetic, spelling, handwriting, and the fundamentals of English respectively. Two chapters dealing with the making of case studies and preparation for remedial teaching, comprising Part 3, complete the book.

Dr. Blair is not concerned with the theoretical implications and ramifications of remedial teaching on the secondary level, including even college, as he suggests. He does not treat remedial teaching on the secondary level as a necessary evil, the need for which ought not to exist and which ought to be prevented. Reading, Dr. Blair states "is such a complex skill that it is possible for the elementary school merely to initiate the process and to develop a few of the basic skills" (p. 3). Accordingly, he devotes his first chapter—the introduction—to statements as to the importance of reading, the extent of reading disability on the secondary level, the effects of reading retardation, and the meaning of remedial teaching. Thereafter, he sets forth the procedures, techniques

and plans involved in remedial teaching, in the faith—while admitting the desirability of thorough preparation—that any competent teacher can make a contribution to this area. In the words of Dr. Blair, "For after all, remedial teaching is just good teaching" (pp. 405-406).

The subjects dealt with range from the means of locating the deficient student and the causes of deficiency and their discovery, to the materials and exercises for the improvement of these deficiencies and the manner of employment of such materials and exercises.

A nation-wide survey of remedial teaching in secondary schools, made by the author in the spring of 1940 and discussed in chapter six of the book, formed one of the chief sources for the book's material.

Twenty illustrations and 22 tables, thorough documentation for practically every important fact and/or statement made, and additional extensive references at the end of each chapter, except the last, enhance the value of the volume.

JACOB I. HARTSTEIN.

*Long Island University
and Yeshiva University.*

DEUTSCH, A. *The mentally ill in America: a history of their care and treatment from colonial times.* New York: Columbia University Press, 1946. Pp. xvii + 530.

This book was originally published in 1937 by Doubleday, Doran and Company. After being out of print for some time, it has been re-issued by Columbia University Press. The book is well documented with eighteen pages of bibliography, arranged by chapters. There is also a well-prepared sixteen-page index.

Tracing the history of the care and treatment of mental illness from earliest historical times to the founding of the American colonies, the author presents the colonial period as one of the most shocking periods of our social history as reflected in attitudes toward the mentally ill. From this period, the author traces the historical evolution of changing attitudes and concepts up to the present-day mental hygiene movement, including the story of Clifford Beers, its founder.

The author describes the end of the American Revolution as the real beginnings of American psychiatry under Dr. Benjamin Rush, "the Father of American Psychiatry." Following him, a period of regression ensued until the middle of the nineteenth century when Dorothea Lynde Dix brought the attention of a shocked nation to the care and treatment of the mentally ill and catalyzed it into constructive action. The next outstanding period was the turn of the twentieth century which heralded "the coming of age of psychiatry in America." At this time "asylums" were beginning to be called "hospitals," with the gradual ascendancy of the therapeutic over the custodial ideal, the growth of private psychiatric practice, and the rise of out-patient clinics.

Integrated with this history are chapters on the rise of State responsibility for the mentally ill, glimpses of the outstanding psychiatrists who have contributed to the field, the epic battle between proponents of restraint and non-restraint, the care and treatment of mental deficiency, the criminally insane, laws governing the commitment of the mentally ill, and a final chapter on mental hygiene. This latter chapter provides a dynamic prospectus for the mental hygiene movement and is well worth quoting for its intrinsic value:

A world of peace and freedom, from which the twin specters of war and insecurity will be banished, a world of equal opportunity, where people will be freed from stunting inhibitions and 'guilt feelings' arising from outworn prejudices and taboos, a world where children may lead healthy, happy lives and grow into useful, well-adjusted citizens, where the personality is permitted to develop naturally and freely, where the individual is given a sense of personal worth and dignity, and where his activities and ambitions are integrated with the development of group life—such is the goal toward which mental hygiene must strive.

Psychologists may well take exception to one serious omission in an otherwise well-organized book. Although he recognizes the role that social work has played in this field, practically no recognition is given to psychology's contributions, except for the section on mental defect wherein the author discusses intelligence test development. The author fails to include the part played by psychologists in assisting psychiatrists in diagnosis and treatment through programs of psychological diagnostic testing, vocational and educational counselling, various types of therapy, clinical research and other related activities.

However, the book is an important contribution to the history of mental disease in this country. It charts the path that must be followed if mental hygiene, as well as psychiatry in its true meaning of mental treatment, are to go forward to greater heights of accomplishment rather than to become static or to regress.

JULES D. HOLZBERG

Connecticut State Hospital

DUNLAP, KNIGHT. *Personal adjustment*. New York: McGraw-Hill, 1946. Pp. xii+446.

Though apparently meant for the undergraduate and the layman, this book appears to promise little enlightenment on the psychodynamics of maladjusted behavior. The treatment is dogmatic rather than scholarly, as evidenced by the author's choice of expressions, his omission of a bibliography, and his neglect to cite the evidence for atypical generalizations. For example, the reviewer would have appreciated evidence for the statement that "most neurotics are vegetarians and have been vegetarians for the greater part of their lives."

Had the author merely failed to make a contribution to knowledge or understanding, the book would hardly deserve mention. However, the book appears to have had a mission, as stated in the preface, where Dunlap writes that "it is high time that psychoanalysis should be presented in its true light." This purpose does not seem to have been fulfilled, since the author has evidently succeeded only in presenting psychoanalysis from a strongly biased point of view. Few contextual opportunities were missed to malign Freud and psychoanalytic concepts. To the reviewer, it seems that it is no longer fashionable among psychologists to regard psychoanalytic concepts as if they were merely the products of "mythology" and "superstition."

Dunlap's first attempt in the present book to present psychoanalysis "in its true light" states: "The progress that was expected a generation ago to be a resultant of co-operation of psychologists and physicians was thwarted by the rise of psychoanalysis, which is based on ancient popular superstitions." It occurs to the reviewer that the prejudice of American psychologists a generation ago, and their lack at that time of techniques by which to deal effectually with personality dynamics, should share in the responsibility for this failure in co-operation. Probably the public expression by psychologists of attitudes such as those expressed by the author will prolong the difficulties in co-operation which the author deplures.

The reviewer fails to see wherein the total effect of this book can be of service to the psychological profession or to the teaching of psychology. At many points Dunlap states atypical opinions as if they were the generally accepted doctrines of psychologists. The author might have profited by the admonition of modern semanticists that one's own opinions should be so identified rather than being attributed to everybody in possession of the relevant evidence.

In fairness to the author, it should be mentioned that his chapters on sex and marriage are vividly frank in the treatment of a subject which often is veiled or avoided in college texts of this sort. Had the remainder of the book maintained the standard set by these chapters, the reviewer might have foreseen for it a more favorable reception than it deserves in its present form.

BERT R. SAPPENFIELD

Montana State University

BAXTER, E. D. *An approach to guidance*. New York: Appleton-Century, 1946. Pp. xii+305.

This book consists of two parts. Part I is a fictional presentation of the experiences of a director of guidance in a small public school. Her activities, problems, and viewpoints are illustrated in specific incidents and conversations with the school administration, teachers, parents,

and others. Part II is entitled *The Story Interpretation* and attempts to state in concise terms the principles and methods of guidance and education illustrated in the story. There are 138 such interpretations, each clearly related to the story proper by marginal references. Throughout both the story and the story interpretation supporting references are made to recent literature and the volume ends with a 227-item annotated bibliography.

With respect to method of presentation, the book is admirably done. The story is interesting and realistic, the interpretations are to the point, and the bibliography is skillfully annotated—a rare achievement in itself. With respect to content, however, the book can be evaluated fairly only if the orientation and purposes of the author are clearly recognized. In the opinion of the reviewer, the author's attitude toward guidance is best expressed in the following quotations: "every teacher is a counselor" (p. 199) and "adjusting, happy teachers mean adjusting, happy pupils" (p. 3). The "guidance expert" will thus look in vain for a detailed treatment of the use of cumulative records, aptitude tests, home-room programs, administrative charts, etc., such as presented by Traxler, Darley, Williamson, Strang, Reed, and others. Instead he will find emphasis upon inter-personal relationships among pupils, teachers, parents, and administration and upon the place of the guidance director in facilitating these relationships. "Guidance" becomes almost synonymous with "education" and the approach to better education is through the teacher's own adjustment and personality growth and the cultivation of a greater awareness of individual needs.

An Approach to Guidance will therefore find its place in the literature on guidance: (1) as a broadening experience to guidance specialists overly engrossed in the mechanical details of administering tests and programs; (2) as a revelation to school administrators of the need for teacher, as well as pupil, guidance; (3) as an inspiration to teachers-in-training; (4) as a general commentary on the goals of education in a democratic society; and (5) as a guide to guidance directors undertaking a new assignment. It is *not* a text book in guidance methods, however, and will serve primarily as supplementary reading in courses on student personnel work, public school guidance and student counseling.

ALBERT S. THOMPSON

Vanderbilt University

BLACK, IRMA S. *Off to a good start*. New York: Harcourt, Brace, 1946. Pp. xii + 256.

In an informal, often humorous way this "handbook for modern parents" discusses the common problems and perplexities that are an integral part of parenthood. Parents will appreciate not only the sen-

sible and non-technical advice directly given but also its underlying assumption that parents are on the whole a well-meaning lot of people doing a difficult job well. It urges parents to accept and respect their own personalities as well as those of their children, and emphasizes the reassuring fact that there are thousands of ways of being successful either for a parent or a child.

The usual range of fundamental problems is covered and many of the aggravating, if often minor, ones also included. The discussion of the essentials of feeding and habit training is rich in practical suggestions as well as basic in modern theory of the psychological significance of these learnings. Discussion of the child developing as a member of the family group includes good material on discipline, sex instruction, the special problem of the precocious child, as well as the minor worries of birthday parties, Santa Claus and the subtle torments of a week of enforced indoor play. Parts III and IV take the child out into the larger world of relations to other children, to adults from Grandpa to dentist, and the widening horizons of travel and schooling, including an analysis of the progressive school's status. The concluding section on intellectual growth and self-expression gives many helpful suggestions concerning play materials and the problems of readiness for the three R's, together with an evaluation of the roles of radio, movies and circuses in the child's life.

MIRIAM FORSTER FIEDLER

Vassar College

BOOKS AND MATERIALS RECEIVED

- ABRAHAMSEN, DAVID. *The mind and death of a genius*. New York: Columbia Univ. Press, 1946. Pp. viii+228.
- AKHILANANDER, SWAMI. *Hindu psychology: its meaning for the West*. New York: Harper, 1946. Pp. viii+241.
- ARTHUR, GRACE. *Tutoring as therapy*. New York: The Commonwealth Fund, 1946. Pp. ix+125.
- BARKER, R. G., WRIGHT, BEATRICE A., AND GONICK, MOLLIE R. *Adjustment to physical handicap: a survey of the social psychology of physique and disability*. Social Science Research Council Bull. 55. New York: Social Science Research Council, 1946. Pp. xi+372.
- BAXTER, DOROTHY. *An approach to guidance*. New York: Appleton-Century, 1946. Pp. xii+205.
- BRIDGES, J. W. *Psychology—normal and abnormal*. (Rev. Ed.) Toronto: Pitman, 1946. Pp. viii+470.
- DELGADO, HONORIO. *La personalidad y el carácter*. Lima, Peru: Talleres Graficos de la Editorial Lumen S. A., 1946. Pp. 207.
- FIEANDT, KAI VON. *Yhteisoelämästä Kansanopiston Luontiosilla Opetuskursseilla*. Jyväskylä, Finland: Weilin & Goös, 1946. Pp. +159.
- JELLINCK, E. M. *Phases in the drinking history of alcoholics—analysis of a survey conducted by the Grapevine, official organ of Alcoholics Anonymous*. Memoirs of the Section of Studies on Alcohol, Yale University. New Haven: Hillhouse Press, 1946. Pp. xi+88.
- KABACK, GOLDIE RUTH. *Vocational personalities—an application of the Rorschach Group Method*. Teachers College, Columbia Univ., Contr. to Educ., No. 924. New York: Bureau of Publications, Teachers College, Columbia Univ., 1946. Pp. x+116.
- KLUCKHOHN, CLYDE, AND LEIGHTON, DOROTHEA. *The Navaho*. Cambridge: Harvard Univ. Press, 1946. Pp. xx+258.
- MASE, D. J. *Etiology of articulatory speech defects*. Teachers College, Columbia Univ., Contr. to Educ., No. 921. New York: Bureau of Publications, Teachers College, Columbia Univ., 1946. Pp. viii+85.
- MAYS, A. B. *The concept of vocational education in the thinking of the general educator, 1845-1945*. Bureau of Educational Research Bull. No. 62. Urbana, Ill.: College of Education, Univ. of Illinois, 1946. Pp. 107.
- PAULSON, BLANCHE. *Wellsprings of action*. Self-Appraisal and Careers Series No. 3. Chicago: Board of Education of the City of Chicago, 1946. Pp. 59.
- REIK, THEODORE. *Ritual: psychoanalytic studies*. (The psychological problems of religion, Vol 1.) New York: Farrar, Straus, 1946. Pp. 367.

SMITH, G. MILTON. *A simplified guide to statistics—for psychology and education*. New York: Rinehart, 1946. Pp. xiv+109.

TRAXLER, A. E., AND TOWNSEND, AGATHA. *Another five years of research in reading: summary and bibliography*. Educational Records Bull. No. 46. New York: Educational Records Bureau, 1946. Pp. vi+192.

WITMER, HELEN L. (Ed.) *Psychiatric interviewing with children*. New York: The Commonwealth Fund, 1946. Pp. v+443.

WOLFF, WERNER. *The personality of the pre-school child*. New York: Grune & Stratton, 1946. Pp. xvi+341.

Forty-first annual report. The Carnegie Foundation for the Advancement of Teaching (1945-1946.) New York: The Carnegie Foundation for the Advancement of Teaching, 1946. Pp. 160.

How to make paper-and-pencil tests. Personnel Research Section, The Adjutant General's Office, Classification and Replacement Branch. New York: The Adjutant General's Office, 1945. Pp. 44.

Personnel classification tests. War Department Technical Manual 12-260. Washington: Government Printing Office, 1946. Pp. iv+90.

Selection and employment of transit operating personnel. A set of five manuals prepared by Glenn U. Cleeton *et al.* New York: American Transit Ass., 1946.

Psychological Bulletin

478 287

CONVULSIVE BEHAVIOR IN THE RAT

FRANK W. FINGER

University of Virginia

I. INTRODUCTION

Convulsive behavior, whether in the human or the infrahuman subject, invariably excites the interest of the observer because of its dramatic deviation from typical response patterns. To contribute significantly to our scientific understanding of behavior, however, the observations of seizures under a variety of conditions must be related systematically to one another, with an attempt made to delineate a limited number of causal sequences common to the several situations. So far as I have been able to ascertain, this has not been done for the laboratory rat. This review is designed to bring together some of the salient features of experiments utilizing this animal, with the hope that the juxtaposition of certain findings will stimulate integrative thought in the form of theories of behavior mechanisms.¹

The independent variable which the majority of investigators have employed to induce seizures in the rat is intense sound. The earlier studies of this type have already been summarized (40). But other experimenters have reported that convulsions can also be produced by such factors as conflict, electrical stimulation, injection with certain drugs, and deprivation of any of several dietary elements.² At first glance these major variables may appear so diverse that their logical organization under a single heading is impossible. But further consideration suggests that the interrelationships among the numerous parameters are so complex that it would be almost as artificial to establish five mutually exclusive categories of seizure. The convulsions observed in the several situations may or may not reduce ultimately to

¹ Convulsive behavior has also been studied experimentally in such mammals as the cat (e.g., 43, 116, 138, 140, 173), the dog (18, 23, 70, 147), the mouse (60, 67, 117), the rabbit (139), and the pig (18).

² There is no reason to limit the list of experimental convulsants to these few. Cobb (20) summarizes 56 clinical causes and 13 physiologic mechanisms that may underlie convulsive behavior. (Cf. also 172.)

a common physiological denominator. But even if there still remain two or more "kinds" of convulsion, an attempted unification at the present stage may well lead to the development of new methods of studying the nature of one or another of these varieties.

The question may legitimately be raised as to the psychological importance of investigations of convulsions which adopt the rat as subject. Enough has been written in defense of the comparative method that little need be added here. It should be stressed, however, that no transfer of principles from rat to human behavior can safely be made until a clearer insight into the complex problem of the rat behavior itself is achieved. For the present, then, this field remains of direct interest primarily to the animal experimenter.

There are three principal reasons why the investigation of induced convulsions is of value.

1. The appearance of the seizure may be influenced by a number of factors—drug injections, dietary deficiencies, the physiological concomitants of aging, chronic emotionality, and conflict, for example. As a consequence, this pattern of behavior might be employed as an assay indicator of the presence and strength of such factors. With few exceptions (e.g., 8), this possibility has been little exploited, probably because of insufficient standardization of the indicator itself. Experiments with this orientation are presumably to be based on such relationships as are discussed in Section IV of this review.

2. The convulsion may alter other behavioral and physiological characteristics of the organism. Several investigators have reported the modification of learned patterns following seizure, and others have attempted to demonstrate changes in diurnal activity, physiological condition, and general reactivity. These experiments have for the most part failed to yield unequivocal results, but continued work in this direction will do much to clarify the neurological bases of these behavior segments and contribute to our fundamental understanding of such matters as the learning process. Section V summarizes the few studies bearing on this type of problem.

3. The nature of the convulsion is of considerable interest for its own sake. Its outward resemblance to human epileptic and induced convulsions has frequently been pointed out, and the similarity of appearance may be paralleled by a similarity of mechanism. If it should eventuate that nothing can be contributed to the understanding of these human syndromes, at least there is every reason to expect that the controlled observation of the rat seizure will expand our knowledge of basic neurophysiological processes. Theories attempting to describe the physiological substrate are a step in this direction; they are considered in Section VI.

II. THE CLASSIFICATION OF CONVULSIONS

From the historical point of view, one might be justified in setting up five categories of convulsions as observed in the rat, each "kind" corresponding to a relatively unique set of experimental conditions.

That this is probably too simple a treatment has already been noted, will be elaborated upon in Section IV, and will be the premise upon which the logical organization of Sections IV and VI will depend. In the present section the oversimplified five-fold classification will be outlined, each classification based upon one major causal factor or kind of factor ("invariable antecedent") that has been proposed by some investigators as accounting for certain convulsions.

Conflict-induced Convulsion ("Experimental Neurosis")

The experiments which gave the initial impetus to the psychologist's interest in rat seizure behavior were designed to demonstrate the efficacy of conflict in the production of behavior aberrations. In 1938 Maier and Glaser (98) reported that rats forced by an air blast to respond to an insoluble problem in a Lashley jumping apparatus sometimes exhibited behavior termed by them "experimental neurosis." The rat might be confronted, for example, by just one card, responses to which had in the past been punished. The resultant neurotic behavior was described (92) as including circular running at terrific speed, during which collision with objects in the room was frequent; tonic and clonic spasms of variable extent and intensity; and a peculiar hopping locomotion. Following this active phase a passive phase occurred, during which the apparently unconscious rat would maintain unusual postures imposed upon it and was extremely insensitive to stimulation. Recovery was gradual, and within a few minutes the rat appeared quite normal, although his cage behavior was described as typically "retiring" (92, p. 30).

In later researches Maier and his collaborators found that the "neurotic" reactions were inhibited by preliminary stimulation with sound, lowering of body temperature, excessive muscular activity (94, 101), the development of "abortive" jumping (93) or position habits (103), and the previous occurrence of attacks (99). On the other hand, subconvulsive doses of metrazol (106), increased intensity of the auditory stimulus (99), and confinement in a small, opaque stimulation chamber (99, 101) tended to increase the convulsive response to the "conflict" engendered by the noisy air blast. Hereditary influences were also demonstrated (95, 100).

The first doubt that this abnormal pattern was dependent upon conflict was raised by the report by Morgan and Morgan (122) that rats lacking discrimination training and stimulated in their home cage by the high-pitched tones of an air blast sometimes exhibited seizure behavior outwardly identical to that of Maier's rats. Indeed, as was

later pointed out (118), every instance of the "neurotic" seizure reported by Maier occurred in the presence of loud noise, and some were observed completely outside the discrimination situation, when the only apparently invariable stimulus was intense sound. Other stimuli, electrical and mechanical, were adopted by Maier to force the rats to jump in the difficult problem situation, but none led to the convulsive pattern. The suggestion was inevitable that it was sound, rather than discrimination-conflict, that precipitated the convulsion. Other investigators (4, 38, 73, 86, 109, 126, 130, 148, 151, 154, 155) have corroborated the observation that untrained animals can often be made to convulse by a few seconds³ of auditory stimulation. In fact much of Maier's later work has been concerned with the production of seizures by various sound sources, with no attempt made to induce conflict by discriminative or other training (cf., e.g., 94, 95, 99, 100, 101, 102). Because sound seems to be the most invariable antecedent of the convulsive seizure, because no conflict between formally learned tendencies is required, and because the changes originally described by Pavlov (137) as characterizing experimental neurosis (temporary loss of discrimination ability, resistance to the experimenter, and alterations in general modes of behavior outside the experimental situation) have not been demonstrated, the majority of investigators now prefer to regard the behavior as something other than conflict-induced experimental neurosis (e.g., 10, 41, 83, 121, 153).

Maier's original position is, however, not undefended. Maier himself states that although "... auditory stimuli are basic in the production of abnormal behavior" (99, p. 29), "The experimental neurosis ... is not a response to air alone" (92, p. 67). The situation involving simple auditory stimulation can be regarded as conflictful, "since the auditory stimulus ... [gives] rise to no specific response but [is] excitatory in nature and the enclosed space restrict[s] general behavior and serve[s] as the inhibitory aspect of the situation" (93, p. 369). Bitterman (12, 13) has argued in the same vein, holding that in the stimulation situation there is conflict between the excitatory auditory stimulation and the inhibitory physical barriers (including empty

³ Smith (153) found that certain peculiar "premonitory" behavior was followed by the full seizure in more than 80% of his 1309 rats. This frequent coincidence would make plausible the inclusion of the premonitory behavior as the first part of the convulsive pattern. Using this as the indicator of onset, many attacks were observed to begin within one second after the beginning of the auditory stimulation. The more obviously convulsive components were observed to have latencies of 1-7 sec. in 50% of his cases (cf. 13, p. 118).

spaces). More direct support of the conflict (vs. auditory) hypothesis is found in the results of two experiments in which sound was apparently not a factor. Griffiths (56) describes in considerable detail convulsions occurring in a rat forced by electric shock to respond in a difficult choice situation, and Bitterman (14) reports one seizure in a learning situation in which shock was the punishing agent. If additional experimental evidence of this nature is forthcoming, and if it cannot be attributed to such an alternative factor as dietary deficiency or electrocortical shock, the existence of the "conflict-induced convulsion" category will become more logically acceptable.

Sound-induced Convulsion ("Audiogenic Seizure")

Although the physiological explanation is still obscure, there is no question but that exposure to intense auditory stimulation will cause a large number of rats to exhibit a pattern of response deviating markedly from typical laboratory behavior (4, 38, 73, 86, 99, 109, 126, 130, 148, 151, 154, 155).⁴ Sound sources successfully used include an air jet, the Galton whistle, buzzers and bells, jingling keys, and an oscillator. No previous training is necessary, nor has any extrinsic factor but sound been demonstrated to be essential. With a latency varying from less than five seconds after the onset of stimulation to somewhat more than one minute, the previously "normal" rat begins a reaction closely reminiscent of the pattern described by Maier as experimental neurosis. The pattern is far from stereotyped, various combinations of wild running, tonic-clonic convulsions, and stupor being possible.

This reaction was probably first mentioned in 1924 by Donaldson (28, p. 134), who, in describing the acute hearing of the white rat, noted the "almost maniacal running and jumping" that often occurred when a bundle of keys was jingled before the home cage. Investigators of the effect of magnesium deficiency in the rat described as early as 1932 a convulsive sequence that occurred "spontaneously" or upon auditory stimulation (51, 85), and Dice (27) reported a similar pattern in "epileptic" mice exposed to the sound of jingling keys. It was not until 1939, however, that this atypical response came to be widely known to comparative psychologists, largely as the result of the controversy over the nature of the Maier-type "experimental neurosis." Because of the apparent primacy of the auditory component in inducing the convulsive reaction, Morgan and Waldman (123) suggested the descriptive phrase

⁴ Sound has also been observed at times to induce various "abnormal" patterns in the human (e.g., "musicogenic epilepsy," 23), the dog (23, 124), the cat (5), the guinea pig (171), and the mouse (27, 117, 168).

"audiogenic seizure," which has become the most widely accepted term.⁵

The audiogenic seizure is influenced in both its pattern and frequency of occurrence by a number of secondary factors, so that susceptibility figures within a colony and in a particular testing situation range from 10% or less (86, 122) to well over 90% (34). The inhibiting effect of the following factors has been demonstrated: strenuous activity before or during the stimulation period (2, 76, 101); extreme restriction of movement (54, 88); the establishment of position habits (103) or the development of abortive jumping (93) when the stimulation is in the Lashley apparatus; the appearance of a fear reaction (2); previous exposure to the convulsion-producing sound (11, 38, 76, 99, 153); electrification of the floor during the auditory stimulation (15); the provision of a "psychological shelter" (56); taming (76); temperature changes (101); chronic "emotionality" (11, 114); vagotomy (88); partial curarization (88); and the administration of such substances as adrenin (2, 87), dilantin sodium (21, 54, 148), adrenal cortex (54), thiamin hydrochloride (59, 81), atropine (72, 73), B-complex vitamins and minerals (131, 133, 134). The frequency or severity of seizure has been found to be increased by such means as repeated anger stimulation (163, 164); increased intensity of auditory stimulation (48, 120, 121); chronic disturbances induced in other situations (75, 76); increased conflict, e.g., as provided by additional barriers to escape (99, 103); cutaneous stimulation furnished by water spray (97); temporary deprivation of food and drink (38); lactation (127); perhaps prenatal treatment with sodium bromide (64); B-complex avitaminosis (133); magnesium deficiency (135); and injection of subconvulsive doses of strychnine (2, 153, 154), metrazol (78, 103, 106, 145, 154), caffeine sodium benzoate (154), sodium phosphate (50), and coramine (166). Age differences (31, 33, 39, 102), sex differences (33), genetic background (34, 39, 57, 58, 95, 100, 113), blood composition (86), inanition (132), cortical insult (7, 153, 169), punishment for seizure (107), and eserine (72, 73) and phenobarbital (77) injections have also been studied with respect to their effect upon audiogenic seizures.

The nature and significance of the sound-induced seizure have been considered variously. The majority of investigators seem to be seeking an explanatory principle involving a rather direct overloading of the nervous system by the auditory stimulation. A few (12, 93) still hold

⁵ Other terms applied to the behavior in question include "audio-epileptic seizure" (153), "epileptoid convulsive reaction" (4), "noise-induced seizure" (72), "noise-fright convulsion" (155), and "sonogenic convulsion" (117).

the seizure to be a manifestation of strong conflict between opposed tendencies, and others (2, 56, 160, 164) emphasize the intense emotional components of the precipitating situation. It may be, as Patton has intimated (129), that the so-called audiogenic seizure is simply a mild form of the disorders which clearly result from severe nutritional deficits. This would appear to be an experimentally fruitful suggestion, the verification of which will probably wait on more refined neurophysiological and biochemical techniques. In any case, the apparently intimate connection with convulsions otherwise produced makes it relatively meaningless at this juncture to consider the audiogenic seizure as an isolated phenomenon.

Electrically-induced Convulsion

Another method of producing convulsive behavior in the rat is the passage of electric current through the skull (Ferrier, 1873, 172). The technique is analogous to human electroshock therapy, and the results are in part similar. The electrodes are variously applied: to the shaven skull (80), one to the head and the other to the tail (125), to the ears (29, 149, 155), to either side of the head between the eyes and the ears (48), or on a line one centimeter behind the eyes (162). The shocks employed to produce the generalized convulsion have included 10-20 volts, 30-50 milliamperes, 0.3 seconds (66); 40 v., 20-25 ma., 2 sec. (80); 70-85 v., 0.2-0.6 sec. (29, 149, 155, 162); 120-145 v., 50-100 ma., 0.1 sec. (174); 600 v., 4.5 ma., 2 sec. (125); and 700 v., 9-14 ma., 0.75-0.95 sec. (48). At these intensities a convulsion consisting of tonic flexion, often succeeded by extension and clonic contractions, occurs in most subjects, usually followed in 10-30 sec. by a period of hyposensitivity and reflex depression. This reaction is attributed by Kessler and Gellhorn (79) to a hyperglycemic condition, in which the autonomic balance is on the sympathetico-adrenal side. Wortis *et al.* (174) suggest that the mechanism involves inhibition of the oxygen uptake of the brain (cf. 25).

Heilbrunn (66) has found that light ether anesthetization attenuates or abolishes completely the electroshock convulsion, and according to Page (125) alcohol tends to raise the shock threshold. On the other hand, atropine sulphate, synthetic vitamin K, calcium gluconate, thromboplastic suspension of brain substance, and combinations of these (66), and adrenalin (125) have been found ineffective in diminishing the effect of the shock.

When a shock of decreased intensity or duration is substituted, the generalized convulsive pattern is sometimes replaced by one resembling

the audiogenic seizure more closely, in that the tonic-clonic phase is preceded by a period of running. This running has been described by Page (125) as being more controlled and "less forced" than that typical of the audiogenic pattern. However, Golub and Morgan (48), Stainbrook (155), and Siegel and Lacey (150) seem to have succeeded in reproducing more exactly the running picture characteristic of the audiogenic convulsion. This production of the pattern attributed by some to conflict or to sound is presented as strong evidence that neither of these factors is as significant in the etiology of the running-convulsive behavior as sometimes implied. Such reasoning is not indisputable but at least suggests that a rapprochement of the differing viewpoints and experimental findings, perhaps at the neurophysiological level, may be mutually enlightening.

Drug-induced Convulsion

As early as 1864 convulsions were produced in experimental animals by the injection of foreign substances (Morcé, 172). With the widespread adoption of drug therapies, it was to be expected that a large number of investigators should turn to the rat in an effort to ascertain the nature and therapeutic mechanism of the convulsions. Metrazol has been the most frequent convulsant studied (e.g., 68, 80, 89, 142). In one investigation (105), a 1% solution of metrazol in 0.85% saline was injected intraperitoneally, with an average convulsive threshold found in one group to be about 40 mg. per kilogram body weight, and the latency of attack approximately three minutes. Tonic flexion, usually with torsion of the spine, is the most characteristic feature of the convulsion, usually followed by a clonic stage. A passive phase terminates the seizure, sometimes preceded by a brief period of rigidity or stiff hopping. Recovery appears complete within a few minutes. The tolerance to the drug has been found to change with repeated administrations, increasing when treatment is given on alternate days, decreasing when injection is weekly (143). The mechanism underlying the seizure has been suggested to involve increased cerebral metabolism and anoxia (25, 174), and increased autonomic (especially sympathetic-adrenal) excitation (37). The discovery that rats maximally susceptible to sound-induced convulsions have on the average a lower drug threshold and a metrazol convulsion of longer latency and duration than do rats less susceptible to sound-induced seizures (105, 144) underlines the necessity of interrelating the patterns appearing in these different experimental situations.

Insulin has also been employed as a convulsant. Berman and Riess

(9, 141) injected one unit per 12 grams body weight, which led in approximately 50% of instances to severe convulsion. Hypoglycemia is very probably a causal factor (8), and inhibition of brain oxygen uptake is also probable (174).

There is no reason to limit the list of convulsants administered to the rat to these two, as any list of pharmacological properties will testify (69). The fact that subconvulsive doses of many chemicals modify the convulsive response to sound or conflict enhances the likelihood that common physiological bases for the different "kinds" of convulsion may be discovered through this type of investigation.

Convulsion Associated with Dietary Deficiency

Rats deprived of one or more elements of the vitamin B complex (17, 131, 132, 134), normal dietary magnesium (26, 124, 135), or calcium and vitamin D (50) become abnormally sensitive to stimulation. If the deprivation is begun early enough these rats will respond to mild auditory or cutaneous stimulation with behavior described as "epileptic" (19) or "uncontrolled" (84)—running fits accompanied or followed by convulsions (24). Concomitant physiological changes are often apparent; Kruse *et al.* (85) found that hyperemia and vasodilatation began about four days after introduction of a magnesium-low diet and subsided 12 days later, with the onset of seizures, death usually intervening within another four weeks. As the deficiency increases, convulsions are often observed to occur spontaneously, i.e., without obvious relation to particular environmental stimuli (52, 135, 136).

The behavioral similarity of the attacks occurring under conditions of deprivation to the "conflict-induced" convulsion and the "audio-genic" seizure might inspire a search for a common physiological etiology. This approach is still more strongly suggested by the sequence of events as deprivation becomes progressively more severe. For example, with a diet including 1-2 mg. of magnesium per 100 g. of dry food, convulsive seizures occur only upon application of a stimulus such as an air blast (51, 52). With only 0.4 mg. of magnesium per 100 g. of dry food, convulsions occur in response to less and less intense sounds as the period of deprivation continues (53, 85), and finally "spontaneous" seizures may be observed (135). Pyridoxine deficiency runs a similar course (136). (It is pertinent to note that sounds are by far the most effective stimuli for releasing these convulsive seizures.) One tentative explanation of the difference between various strains of rats in susceptibility to sound has been a difference of diet, which becomes an obvious possibility in light of the above studies. Patton (129, 136) by implication

goes further. With a mildly deficient diet he obtains approximately 100% convulsive responses to sound, and with more severe deficiency "spontaneous" seizures in nearly all rats; with a diet fortified by mineral and vitamin supplements the frequency of "audiogenic seizure" is close to zero. It is no great extension of this relation to propose that underlying all sound-induced convulsions is a dietary deficiency, and that the Maier-type "neurosis" is no more conflict-induced than were the symptoms attributed by Kruse *et al.* (85) in 1932 to magnesium deficiency. Until these possibilities are eliminated, it will be well to consider the general topic of "convulsions" as a whole, rather than by categories, evaluating each of the many factors that may modify the behavior (cf. Section IV).

III. THE CONVULSIVE PATTERNS

Before a decision can be reached as to the unity or diversity of the "types" of rat convulsions, it is obviously necessary to compare the patterns appearing in the several situations. It must be confessed that no very confident comparison can be made at present since because of the inherent variability of the patterns, or deviations reflecting testing differences, or inaccuracies of the observations, there is imperfect agreement as to the precise form of behavior that characterizes each "type." Whether or not differences between the categories are reliably greater than differences within a single category, outweighing the similarities, probably will be settled only on the basis of refined photographic techniques.

Convulsion Associated with Auditory Stimulation. Whether the more significant causal factor is sound or conflict, it is agreed that "neurotic" or "emotional" convulsions in the absence of intense sound stimulation have been described in only one rat (56), and this single description offers little detail. Further, the proponents of the neurosis hypothesis seem to impute "conflict" to any situation which includes inescapable stimulation of high intensity (e.g., 13, 93). It is therefore logical to combine under one heading the descriptions of the seizures attributed by their observers to either variable.

Just how complete a pattern of behavior should be included as audiogenic or neurotic convulsion is open to question. Maier (92) and the Morgans (122) originally incorporated the early "blind" running into their description, probably because of its unusual nature and because no true convulsion ever appeared without a preceding running phase. Most writers have followed this practice, although the question immediately arises as to the meaning of the typical running pattern when it occurs alone or, for that matter, of the so-called premonitory behavior

or the substitute activity that so frequently precedes or takes the place of running. The criterion of attack will obviously require careful definition before a truly satisfactory conception of the problem can evolve; numerous controversial issues have remained unsettled largely due to an unrecognized disagreement as to terminology.

Accepting the most complete sequence as constituting the convulsive seizure, we find almost innumerable combinations of the premonitory, running, tonic, clonic, and passive phases. (For detailed descriptions of variations, cf. 7, 92, 99, 119, 153, 157.) Classifications have been suggested for certain combinations of symptoms (99, 157), but the reality probably approaches a continuum (or several continua) of severity and duration. Any generalized description, such as follows, is thus misleadingly simple, although possessing the merit of facilitating comparison with the convulsions evoked by electricity, drugs, and dietary deficit.

After the initial startle response to the onset of the stimulus, any of a number of patterns may appear. Some animals appear to ignore the sound source, others seek to retreat from it, a few actively attack it. Many rats crouch motionless or try to burrow beneath the floor covering or their fellow subjects. In most nonseizure trials some sort of "substitute behavior" appears—rapid nose or ear rubbing, teeth chattering, chewing, shivering, shaking, vibrissae twitching, body cleaning, or restless head and body movements (2, 88). In trials culminating in a seizure, behavior described by Smith (153) as a "motor aura" usually occurs, typically consisting of brief quick runs, jerky sidling or backward steps, or pivoting movements of the head and body (cf. also 63, 114, 151). After a latency ranging between one second and one minute (4, 7, 119, 153), the active phase begins with a running period, sometimes of gradually accelerating speed, sometimes of a very explosive nature, and in any instance finally attaining an unusually high velocity. The running may or may not follow a circular path, but almost always seems undirected with respect to obstacles such as furniture and other rats. This phase may be interrupted by one or more quiet periods, may terminate suddenly in what resembles coma, or may continue with diminishing vigor until exhaustion. Most characteristically, it leads directly (after an average of 15–20 sec.) into the convulsion proper. This may be predominantly of tonic extension, reminiscent of decerebrate rigidity (99, 114), of incoordinated clonic beating of the limbs, or may combine both tonic and clonic aspects. During the tonic-clonic phase (average total duration 30–40 sec.) the following are sometimes observed: vocalization, exophthalmus, excessive salivation, cyanosis, ejaculation, diminished respiration, incontinence, dilated pupils, piloerection, retraction of testes, relaxation of the vaginal opening, and peripheral vasodilatation, with death intervening in a very few cases. The rat may continue to exhibit clonic jerks (e.g., spasmodic jumping) with decreasing frequency as long as stimulation continues, or may lapse into the

phase usually described as comatose. Most animals are very unresponsive to stimulation during this passive phase, with reflexes depressed to varying degrees. Spontaneous activity is virtually abolished, and the animal is nonresistant to handling. This condition may last for longer than 15 minutes, although gradual recovery of spontaneous activity usually begins within two or three minutes. (The decreased activity is probably not simply the result of fatigue, for as Beach and Weaver (7) point out, the duration of the coma is often inversely related to the amount of exertion occurring in the active phase.) In a few instances hypersensitivity or renewed clonic activity succeeds or takes the place of the typical stuporous condition (63, 117, 153). A recovering animal appears extremely fatigued, and quantitative recording reveals that the frequency of movement is subnormal for several hours (42). For some hours following a seizure the threshold to induced convulsion is markedly elevated.

Accompanying this pattern of overt behavior, a number of internal changes have been detected. Lindsley, Finger and Henry (88) demonstrated alterations in the electrocortical activity paralleling the external manifestations of the attack, and although technical difficulties obscured detailed analysis, many features in common with the EEG in human convulsive seizures could be recognized. The heart rate also reflects the progress of the seizure, fluctuating markedly just before the attack, reaching a high level during the active phases of the convulsion, and dropping far below normal and showing various irregularities during the comatose stage (47, 76, 88, 95, 109, 111, 112). The passive phase is characterized also by acidosis, decreased blood CO_2 , and decreased clotting time of the blood (22).

The only attempt to distinguish the pattern occurring in response to sound alone from that resulting from auditory stimulation in a formal conflict situation is that of Maier and Glaser (99). They suggest that in the former situation the extension component of the convulsion is relatively more intense and the passive phase less pronounced. However, these differences could easily be attributed to uncontrolled variables, especially age, differentiating the two groups compared, and Maier and Glaser emphasize that "the similarities are too great to consider these differences as symptoms which distinguish between different abnormalities" (99, p. 12).

Electrically-induced Convulsion. There are at least two different forms of seizure resulting from electrical stimulation of the brain, depending on the intensity of the current. At higher intensities, the onset of the current is marked by a "start," with the spine flexed inward, hind legs drawn up, and front legs brought down (125). For the duration of the shock (conventionally 2 sec. or less) the animal remains rigid. He may then pass into a state of tonic rigidity, with the back arched, the tail stiff, the legs sometimes briefly flexed (79) and then extended

as far down as possible, and the eyes tightly closed. Gradually the fine tremors of this stage give way to clonic movements (involving principally the hind legs, 48, 155) of decreasing frequency and increasing amplitude, with frequent urination, defecation, and sometimes ejaculation. After about 30 sec. coma supervenes. This varies in character, but usually begins with a very deep phase during which the rat is non-resistant and the reflexes abolished even more completely than in the sound-induced seizure. After a few minutes there often succeeds a hyperexcitable stage during which "blind rage" reactions can be observed.⁶ As measured by maze activity, the recovery period is longer than in the sound-induced seizure (160). As soon as the animal begins to be active, a second convulsion can be elicited, in contrast to the longer refractory phase following the audiogenic seizure.

With less intense shock, the tonic phase may be omitted, with the clonic and comatose phases following directly the primary jerk (48). The more characteristic pattern, and of particular interest because of its similarity to the sound-induced seizure, includes a vigorous running phase after the initial "start," which may alternate with or be followed by tonic rigidity and/or clonic spasms before the coma begins. Golub and Morgan (48), Siegel and Lacey (150), and Stainbrook (155, 157) describe this seizure as resembling the audiogenic convulsion closely, although the latter observer notes a greater resistance to handling and more promptly returning activity in the postseizure state. On the other hand, Page (125) believes that the running that is induced electrically appears less "forced" than that following auditory stimulation.

Physiological changes accompanying the electroshock convulsion include a relative cerebral anoxia, attributed to increased cerebral metabolism (25), and extensive hemorrhages in the pia-arachnoid and in the parenchyma (66).

Drug-induced Convulsion. There are a great many drugs which, injected with the proper dosage, produce convulsions in the rat (69). Because of the interest in the insulin and metrazol therapies for human behavior disorders, most of the comparative psychologist's attention has been directed toward these convulsants. For the most part, however, the emphasis has been more upon the effects of these drugs on learning and retention than upon the convulsive patterns themselves.

Berman and Riess (9) injected 1 unit of insulin per 12 g. body weight. After about 40 minutes, lethargy and/or coma were observed. In nearly 50% of the subjects severe convulsions were also observed, involving especially the head region. Recovery appeared complete at the end of 8 hours.

Maier and Sacks (105, 144) give a more detailed analysis of the

⁶ Cf. the "sham rage" reactions of the decorticate dog and cat (5). It has been suggested that the convulsive seizure may sometimes bring about a sort of temporary "functional decortication."

metrazol reaction. The tonic spasm is the most characteristic sign of the syndrome, but two different patterns distinguish rats minimally and maximally susceptible to sound-induced seizures. In the first group, intraperitoneal injections are followed in 2-4 min. by a few twitches of the head and a sudden tonic flexor spasm involving the head and all four extremities, usually with spinal torsion. This is replaced by clonic movements of the forepaws, with the combined tonic-clonic period lasting perhaps one minute. The rat then becomes relaxed and passive to handling (slow pulse, 47), with recovery apparently complete within a few minutes. In the group that responds convulsively to sound, the onset of metrazol convulsion is frequently delayed for several minutes. The tonic flexor spasm and forepaw clonus are usually present, but these stages are followed by a period of rigidity, during which the rat is extremely hypersensitive to stimulation. There then usually occurs a period of rigid hopping (as seen in the "neurotic" rats described by Maier, 92), sometimes leading to a second period of rigidity and repeated convulsions. In both groups of animals it appears that repeated drug injections or, to a lesser extent, sound-induced convulsions, lead to an elaboration of the pattern as well as a decreased threshold of convulsive dosage.

Convulsion Accompanying Dietary Deficiency. Most of the studies describing convulsions in nutritionally deficient animals have utilized diets low in magnesium (53, 85, 124, 136) or certain elements of the B-vitamin complex (17, 24, 84, 134). The two patterns, whether precipitated by auditory stimulation or occurring "spontaneously," seem too similar to allow any clear differentiation. The patterns both typically begin with high speed racing fits, often describing a circular path, and frequently punctuated by vocalization. The rat may then fall on his side for a brief period, rigid and with breathing suspended. Tonic spasms and clonic twitchings then usually occur. Kruse, Orent and McCollum (85) report no incontinence in magnesium tetany; Chick, El Sadr and Worden (17) observed both urination and vomiting during convulsions in rats deprived of B₆, which coincides more closely with the descriptions of sound-induced convulsions. Sometimes only the running aspect is observed (17) and sometimes multiple seizures take place (65, 85). Usually a coma follows the active phase, marked by greatly depressed reflexes and, as in audiogenic convulsions, a slow and irregular pulse (17, 53). The total pattern, active and passive, usually lasts 2-4 minutes, with the gradual recovery coming to the head region first. Death terminates the seizure more frequently than is typical of the other convulsive states.

As Patton, Karn and Longenecker (134) point out, these seizures can be remarkably similar in outward appearance to sound-induced convulsions. To what extent the similarity extends to the physiological level has yet to be determined.

On the basis of the best descriptions at hand, it would appear that the five "types" of convulsion have much in common, as far as outward manifestations are concerned. There is no reason to distinguish between the convulsions precipitated by sound and those (if any) based on sound-induced conflict. While the pattern of response to high intensity electric current lacks the running phase characteristic of sound- or conflict-induced seizure, lower intensities do produce the running-tonic-clonic-coma sequence. Although it is not clear that these sequences are identical, at least similarities are striking enough to inspire comparison at the neurophysiological level. Similarly, the metrazol shock includes components suggestively reminiscent of the audiogenic seizure, although lacking the running phase. Finally, the descriptions of audiogenic and electrogenic seizures are virtually duplicated by those of attacks occurring at certain stages of dietary deficiency. There is little evidence in the overt picture to require a differentiation of this nutritional syndrome from the seizures produced by auditory or electrical stimulation in the presumably well-fed rat.

The external resemblances do not of course establish the community of physiological or psychological mechanisms. The results of more precise investigation may invest apparently slight divergencies with increased significance. Until then, however, we may well orient our experiments around complex patterns of variables which influence convulsions in general, rather than around the more limited number of etiologically designated categories.

IV. VARIABLES INFLUENCING CONVULSIVE BEHAVIOR

The resemblance of the convulsive symptoms induced under the several sets of circumstances is not the only hint that a consideration of convulsive-behavior-in-general may be fruitful. Several investigations make very clear that the appearance of one "kind" of convulsion in a given rat may be in part a function of experience with another "kind" of convulsion, that what appears to be a major factor in one "kind" is a contributing factor in a second "kind" and vice versa—in short, that the interrelationships among the numerous variables are so complex that the intercategory lines are crossed too often to allow their strictly divisional meanings to be maintained. A few examples will clarify the point.

1. The amount of metrazol required to produce convulsion, as well as the precise form of the seizure obtained, is a function of the susceptibility of the rat to sound-induced seizure (105). Further, both the threshold and form of metrazol convulsion are altered by the experience of sound-induced seizures (144). On the other hand, subconvulsive doses of metrazol increase the likeli-

hood of sound-induced seizure, even though such attack still possesses the unmistakable characteristics of sound-seizures (78, 103, 106, 145, 154).

2. The percentage of a population susceptible to audiogenic seizures varies with diet (50, 133, 135). While convulsions in nutritionally deficient rats sometimes appear "spontaneously," sound stimulation releases them most readily (53, 134).

3. According to some investigators, the convulsive reaction to sound is increased if conflict based on difficult discrimination is added (99) or if the animals have been brought to a condition of chronic emotional disturbance (75). Conversely, sound stimulation has been found to be a peculiarly effective means of inducing conflict that may lead to seizure (99).

No behavior pattern can be said to be the result of a single causal factor. Every response is the product of an interaction between the complex environment and an organism whose immediate state has been influenced by a multitude of conditions. It is thus a questionable procedure to assign the responsibility for any reaction or type of reaction to one characteristic of the environmental-organismic situation. Nor is it invariably possible so to manipulate the relevant variables that even the major antecedent can be segregated. In the problem under consideration, for example, it seems that according to one personal definition of "conflict" it is virtually impossible ever to provide an intense and continuing stimulus without concurrently inducing conflict (13). At the same time, the studies of Patton and his collaborators (127-136) make it clear that the dietary background of a rat can never be ignored; it is difficult to know with certainty that the diet of a particular animal is adequate, that with improved nutrition the situation presently leading to convulsion might not lose its effectiveness. Likewise, it is not easy to limit the exteroceptive stimulation impinging on the possibly malnourished subject to a degree that it can be ruled out as a precipitant of seizures. For these reasons, as well as for those already stated, it seems defensible at present to avoid dogmatic etiological classification of convulsions, and simply to outline the results of variation of a number of parameters, hoping that this form of organization may lead to new insights regarding the physiological (or psychological) bases of the behavior.

Auditory Stimulation

With the exception of experiments involving the use of electroshock or drugs, the vast majority of situations producing convulsions in the rat have included some form of relatively intense auditory stimulation. This is true even of certain investigations designed ostensibly to demonstrate the role of conflict (e.g., 92) or dietary deficiency (e.g., 165) in the production of seizures. In the early studies it was not clear which physi-

cal characteristic of the sound stimulus was primarily relevant. Maier and Glaser (99), finding jingling keys a more effective stimulus than the ringing of an electric bell, an air blast, or high pitches produced by a Galton whistle, even though to the experimenter's ear it was the least loud, concluded that "the degree of effectiveness depends more on the quality of the noise than upon the intensity" (99, p. 28). Maier (95) adds later that the optimal quality also varies from colony to colony. On the other hand, it has been shown that with sound frequency held constant, the number of seizures increases with increased intensity, while with intensity unchanging a positive relationship obtains between sound frequency and number of seizures (121). Carrying the analysis further, Morgan and Galambos (120) have demonstrated that the effectiveness of an auditory stimulus depends not upon its absolute intensity level, but upon the sensation level (i.e., decibels above the rat's threshold for the frequency involved), and that even low pitches, if of sufficient loudness, can produce seizures. It is therefore reasonable to presume that the greater effectiveness of the jingling keys was dependent upon the presence of overtones of high frequency, inaudible to the human ear but falling within the range of maximal sensitivity of the rat (cf. 49 for the rat's audibility curve). It is unfortunate that so few investigators have attempted any objective analysis of the sound pattern employed, but instead have depended for calibration upon the "experimenter's ears" (97, p. 285). This reduces the meaning of such statements as "the water test offered much less auditory stimulation than did the key jingling or air hiss tests" (97, p. 285), and vitiates the conclusion that in such instances the auditory component is a negligible factor. Similarly, it may be the uncontrolled variation of loudness, rather than increased conflict, that accounts for increases in frequency of seizure when the subject is tested in a discrimination situation (92) or in an enclosed testing chamber (99). It would seem necessary to assume for the present that the effect of sound depends predominantly upon its sensation level (for the rat), and that this characteristic must be controlled before less measurable aspects of the situation are considered causal.

It must be added that the intensity required to precipitate a seizure varies from individual to individual, and within the same subject from time to time. Greenberg and Tufts (53) point out that as the rat becomes more deficient in magnesium, a less intense auditory stimulus is required to set off the convulsion, and others have found similar results with subconvulsive injections of metrazol (106). It may be expected that a number of factors which influence the reactivity of the receptor and nervous systems will vary the effectiveness of a constant auditory stimulus (cf. many of the following subsections, and 75, 92, 99, 122, 123, 127-136).

The total energy of the auditory stimulus is not always related to

the likelihood of seizure. The number of convulsions is no greater with continuous sound than with alternating 1-second periods of sound and silence, although with longer cycles the interrupted tone becomes less effective (45). It may be significant from the standpoint of threshold mechanisms that the latency of attack is related neither to intensity (119) nor to temporal pattern of stimulation.

There has been no experimental demonstration that it is stimulation of the cochlear receptors that is a determiner of convulsive behavior. Because of the intimate anatomical relation between vestibular and auditory receptors, it is not impossible that the vestibular function is the more important here. This possibility is perhaps strengthened by the finding that swinging the rat during sound stimulation increases the number of seizures (75), by the frequent observation of peculiar circular or pivoting movements just prior to the attack proper and the occasional circular pattern of running, and by the depression of the righting reflex in the postconvulsive stage. Then, too, it may be the action of cutaneous or pain receptors, particularly in the middle ear, that is the really significant feature of the "auditory" stimulation. There is need for a more clear delimitation of the necessary stimulation modality, probably by pharmacological or operative techniques.

Electric Shock

Electric current has been used in several types of rat problem, sometimes as a motivator, sometimes as a punishment or pain-producing agent, sometimes to bring about specific physiological effects. The behavior evoked depends upon the intensity and duration of the shock, the point of application, and the condition of the animal. To yield a convulsive response one may, for example, pass a current of 600 v. and 4.5 ma. through the skull for an interval of 2.0 sec. (for other combinations cf. *supra*, Section II). Such stimulation usually leads to a tonic-clonic seizure of perhaps 30 seconds duration (125), followed by a comatose stage from which recovery gradually takes place over a period of several minutes. With a less intense shock, the tonic-clonic phase is often preceded by the disorganized running that characterizes the sound-induced seizure.

Heilbrunn (66) attempted to abolish the electroshock convulsion by drug injection just before stimulation. Results were negative with the following: atropine sulphate; synthetic vitamin K; calcium gluconate; thromboplastic suspension of brain substance, with or without calcium gluconate; synthetic vitamin K and the suspension of brain substance; and vitamin K, brain substance, and calcium gluconate. He was able to diminish the severity of the seizures, or to abolish them completely, by light ether anesthetic.

When the shock is applied to the rat via other pathways, such as the legs, convulsions do not normally result (6, 10, 15), although extreme

fatigue and passivity may supervene with continued stimulation (30) and fear and resistance to the situation may accompany repeated trials (29). When shock applied to the feet has been used to force the rat to respond in a difficult discrimination situation, convulsions have virtually never occurred (10, 56, 92, 123). Only Griffiths (56) has been able to produce convulsions in such a situation (one of a group of 10 rats), which, it has been suggested by Stainbrook (157), might have been actually the result of cerebral electroshock. The fact that mild peripheral shock is one of several stimuli which have been successfully applied to evoke convulsions in animals with a variety of dietary deficiencies (50; but also cf. 135) suggests an alternative explanation of Griffiths' findings.

Electric shock has been coupled with auditory stimulation in several experiments. Page (125) found that the convulsive threshold to cerebral electroshock was not altered when the rats were tested "while they were exposed to the sound of air blasts" (p. 186). Maier and Glaser (101) electrified the floor of the test chamber every 2 seconds while susceptible rats were being stimulated by the sound of jingling keys. The number of responding animals and the total number of seizures were essentially the same as for control periods of sound stimulation alone. Morgan and Waldman (123) likewise found that the addition of shock to a sound plus conflict situation had no appreciable effect on the number of seizures. The results of Bitterman and Warden (15) seem to contradict these observations, for the distraction of the shock decreased the number of seizures exhibited by their rats. However, the results of their control trials reduce the significance of these findings.

Other Stimuli

There has apparently been no demonstration that other exteroceptive stimuli are capable of eliciting convulsive behavior, except in conjunction with certain factors of an atypical nature. Unusual sensitivity to noise is reported by Chick, El Sadr, and Worden (17) in rats reared on a diet low in vitamin B₆: "the slightest external stimulus seemed to precipitate" seizures quite similar to audiogenic seizures (p. 579). In some of these rats immunity to noise developed, but seizures could still be induced by stroking, by pricking the tail, or by placing the animal in an unusual position. Similar observations are reported by Patton, Karn and Longenecker (134). J. R. Hamilton (65) describes convulsions lacking the running component that have appeared over a period of several years in one laboratory strain of rats. Noises may precipitate the attacks, but tactile stimulation is most effective. On the other hand, stimulation of an intense nature inhibits these seizures.

Other forms of stimulation have been found to affect the capacity of sound to induce seizures. Humphrey and Marcuse (75, 76) present

evidence that swinging or rotating the animal during sound stimulation may increase its susceptibility to seizure, and Maier and Feldman (97) have demonstrated that a water spray has a similar effect. In the opposite direction, immersion in warm or cold water may reduce the number of sound-induced convulsions (101), although the significance of this experiment is obscured by concomitant exercise of a strenuous nature. Stimulation by bright light is also said to inhibit seizures (92). Whether the modifying effect of these several stimuli is mediated by the production of changed emotional status or operates in a more direct fashion cannot be decided on the basis of the experimental data now available.

Age Differences

With increasing age, a number of vital physiological alterations take place in the organism. Which of these changes modify the responsiveness of the rat to sound-induced seizures is not known,⁷ yet there is no question but that the degree of susceptibility is in part a function of age (27, 31, 39, 102, 115). The number of responding rats in a colony seems to increase during the first months of life, reaching a maximum at about 135-150 days and declining thereafter. There are, however, individual exceptions to this trend, with some rats (maintained on standard colony diet) first exhibiting convulsive behavior at 250 days or later. Although a decrease in susceptibility is the normal accompaniment of aging beyond about 150 days,⁸ some animals continue to demonstrate the full pattern at 500-600 days. The question of the relation between severity of attack and age is not clearly settled. In younger rats the seizures seem to be principally of the running type, with the more severe convulsions appearing more frequently with age increment (34, 130). Smith (153), however, reports more deaths during seizure in the younger subjects. According to Maier and Glaser (99), rats over seven months of age have fewer attacks of the "extensor type" than do younger rats.

The age of the rat at the time when withdrawal from or supplement to the diet is made is of importance in influencing the occurrence of "sound-induced or "spontaneous" seizures (cf. *Nutritional Factors* below). A diet lacking in pyridoxine may fail to produce convulsions in a mother, while the offspring will exhibit severe disorders during the nursing period (134). Too, it appears likely that the earlier in develop-

⁷ Alterations in the rate of metabolic processes, the changing endocrine balance, the number of intestinal flora capable of synthesizing vitamin B (134), the atypical heart rate development of the rat (111), and the variations in auditory sensitivity might be relevant.

⁸ Farris and Yeakel (33) trace a decline in the susceptibility of albinos beginning as early as 60 days. The exact time relations will undoubtedly vary from investigation to investigation, depending on such uncontrolled factors as diet, strain of rat, mode of testing, etc. For example, the same experimenters (34) find that in gray Norway rats the percentage of attacks increases up to the age of 80-100 days.

ment a supplement of vitamin B-complex concentrate is given, the more complete the protection against sound-induced seizure (128).

Inasmuch as susceptibility has been shown to vary markedly with age, and in a highly individual manner, it is essential that particular effort be made to control this factor in investigations attempting to ascertain the effect of other variables upon seizures, especially where a wide age range is to be covered.

Sex Differences

There is little clear-cut evidence concerning sex differences in susceptibility to convulsion. Morgan (119) found no reliable difference in the average latency of audiogenic seizure between the two sexes, and no statistically significant differences in over-all incidence has been reported (e.g., 7). But in spite of the lack of conclusive data in any single study, the trend of all available figures is in the same direction—slightly more males than females respond convulsively during auditory stimulation. Unpublished data of the writer show 18.7% of 268 unselected males and 11.6% of 112 females susceptible over a series of four tests, Maier and Glaser (102) report 55.7% vs. 52.3%, and Farris and Yeakel (34) and Finger (39) found differences of a similar order. Farris and Yeakel (33) suggest that males may react more frequently early in life, with responses of females predominating later. Although females might logically be expected to be more variable in their response because of their periodic physiological changes, it is impossible without further experimental data to assign greater total susceptibility to either sex.

Differences in Genetic Background

Strain differences. A number of workers have noticed differences in convulsive responsiveness among rats of different strains. It seems quite clear that domesticated gray Norway rats are much different from albinos in their response to sound (31, 115). In the first place, their startle is appreciably greater (36). Then, the number of animals susceptible is significantly higher, in one instance exceeding 95% while a comparable group of albinos yielded less than 70% of susceptibles (34). The average percentage of trials per rat giving seizures is higher for the gray Norways, the increase with age continuing longer and the subsequent decrease postponed several weeks. Finally, more seizures among the grays, both relatively and absolutely, include the convulsive phase, as contrasted to the predominance of running among the albinos tested with the same auditory stimulus.

The results with wild gray rats are decidedly in contrast. Griffiths (58) was not able to induce a single seizure in 126 wild Norway and 15 wild Alexandrine rats, using air blast, jingling keys, and Galton whistle successively as sources of sound. Unless these negative results can be attributed to an insufficient intensity of stimulation, we must

explain the difference between the responses of the wild and domesticated groups in terms of such uncontrolled factors as age, dietary background, and emotional state.

Differences in inherited "emotionality." A second genetic difference influencing susceptibility to sound relates to Hall's concept of emotionality (61). His primary measure of emotionality is the amount of defecation and urination displayed by the rat when placed in an unfamiliar open field of 8-foot diameter. The tendency to react in this emotional fashion seems to be inherited (62), and rats bred from a so-called emotional strain display fewer convulsive attacks in response to sound than do rats from a nonemotional strain (11, 113, 114; cf. also *infra*, p. 223).

Differences in familial susceptibility. The greater part of the evidence points to the conclusion that, with other factors held constant, rats with susceptible parents are more susceptible, on the average, than are offspring of nonsusceptibles. For example, by selective breeding over a number of generations, one strain can be produced with a relatively higher degree of susceptibility, a second with a much smaller percentage of convulsions (57). The precise mechanism of genetic transmission is still in doubt, however. Dice (27) and Watson (168) concluded that an "epileptic" pattern in mice, apparently somewhat similar to the sound-induced seizure in rats, was transmitted as a Mendelian recessive trait. Maier and Glaser's early study (100) concluded that the susceptibility to convulsion behaved as a simple Mendelian dominant, but the data of Maier's later experiment "do not support a single Mendelian type of transmission" (95, p. 331). The entire hereditary background of a particular rat must be considered in predicting his susceptibility: 68.6% of five litters with nonsusceptible parents and grandparents, but from an "unstable strain," were found to convulse to sound (95; cf. also 57). The source of much of the confusion concerning this matter arises from the neglect of other relevant factors, such as age. As has clearly been shown, the diagnosis of "nonsusceptible" made on the basis of tests over a limited span of life is misleading, for a rat failing to respond at the age of 20 weeks may respond for the first time at 40 weeks (39). Certainly the question of genetic transmission will not be settled until more investigations are made, with systematic control of the other pertinent variables.

A little information is available indicating that convulsive response to metrazol is related to the inherited susceptibility to sound alone (95). This may be further evidence of the unity of the several "kinds" of convulsion in the rat; additional study may reveal the inheritance of a constitution with a certain level of general convulsion-tendency.

Differences in "nervous tone." In one group of rats having experienced a number of audiogenic seizures, the systolic blood pressure was found to be abnormally high (115). Inasmuch as the majority of these animals were from a single litter, the experimenters hesitated to ascribe

the hypertension to the occurrence of seizures. Rather, "audiogenic seizures may be simply a measure of that factor in the animal's constitution ('nervous tone') which causes the hypertension" (p. 303).

Differences in "Emotionality"

As indicated above, rats from a strain bred for "emotionality" (as measured by response in the open field test) are less susceptible to convulsion when exposed to sound stimulation than are rats from a non-emotional strain. When the rats themselves are tested for emotionality in the open field, a higher percentage of those classified as nonemotional are convulsers, the correlation between the two patterns of behavior being $-.57$ (11, 113). It has been suggested that the chronically emotional animals have an habitual "safety valve" which allows the constant dissipation of tension rather than its building up to the threshold of convulsion (114). Requiring a somewhat different interpretation is the discovery that "gentled" rats have fewer sound-induced seizures than do untamed rats (76, 110).

Conflict and Chronic Disturbance

It would be difficult indeed to define "conflict" to the satisfaction of all psychologists and in such a manner that there would be unanimity as to the conflict value of a particular situation. Eschewing this task, I shall simply quote the word as it is used in certain experimental reports; some readers will desire to insert the modifier "alleged" at several points. The controversial literature may shed additional light on the semantics of the situation (e.g., 13, 40, 41, 56, 92, 96, 118, 123).

It has already been noted that a few investigators impute conflict to every situation in which an intense auditory stimulus evokes convulsive behavior (e.g., 13). From this standpoint, the material reviewed above under *Auditory Stimulation* becomes immediately relevant, with the substitution of "conflict" for "sound." Beyond that, there are a number of examples in which a factor termed "conflict" has been examined as to its effect upon convulsive behavior. It would be very convenient for the student of conflict, incidentally, if sound-induced seizures *were* demonstrably independent of conflict. In such an instance, the facilitating influence of deliberately provoked conflict upon a subthreshold convulsive tendency (with a standard auditory stimulus) could become a measure of the degree of the conflict itself. The sound-stimulated rat would then become an assay object for conflict just as it might for an inhibiting or facilitating drug. And similarly for several others of the factors considered in Section IV.

As has been pointed out, Griffiths (56) describes convulsions in a rat forced by an electric current to choose between two negative cards in a Lashley jumping apparatus. The dominant factor as explained by the experimenter is stimulation of the emotional centers, with "Fear and

escape . . . the basic principles underlying the seizures" (p. 29). The proponents of the conflict hypothesis on the other hand could describe the crucial factor as being conflict between the necessity of reacting and the unavailability of an appropriate response.⁹ No other example of convulsion resulting from non-auditory conflict is in the literature.

Conflict brought about by the presentation of a difficult discrimination problem has several times been reported at least to increase the likelihood of seizure in response to sound. Maier and his collaborators (92, 99, 103) found that a greater number of seizures result when the animal is faced with a one-window situation in the Lashley jumping apparatus and at the same time stimulated by an air blast than when stimulated by air blast alone. The greater number of seizures is attributed to the addition of the conflict based on the jumping problem. (On the other hand, it is interesting to note that in the conflict-plus-sound situation the percentage of *convulsions*, contrasted to simple running attacks, may be smaller than in response to sound alone.) Using a small number of rats in a similar experiment, Morgan and Waldman (123) obtained negative results.

Other features of the testing situation are said to add to the conflict and thereby increase the number of seizures in response to sound. Erecting walls on the testing platform produces more convulsive seizures, and enclosing the subject on all six sides leads to an additional increment (99). Whether the greater effectiveness of this arrangement is due to alterations in the auditory component or in the conflict aspect is still debated. When a rat stimulated by sound is simultaneously stimulated by a powerful water spray, more convulsions occur than when the non-auditory effect of the spray is lacking (97). These results have been interpreted as meaning that conflict between the need to escape the annoying spray and the inability to do so increases susceptibility to convulsion. That the facilitating effect of the water depends on more than its cutaneous stimulation value is yet to be established.

Two other experiments may be interpreted as supporting the contention that the incidence of sound-induced convulsions can be increased by conflict. Bernhardt, Tobin and Signori (10) were unable to produce convulsions in a difficult Lashley discrimination situation even though response was forced by electric shock. When rats exposed to this problem were subsequently tested with a loud bell, however, they exhibited more seizures than did a control group lacking discrimination training. Humphrey and Marcuse (75, 76) produced "chronic disturbances" in rats by noisily moving the bottomless feeding box after completion of maze trials. The animals so treated convulsed more readily in response to noise than did a control group. Small numbers and incomplete controls diminish the significance of these experiments, however.

⁹ Alternative explanations are mentioned above.

Anger and Fear Reactions

Tseng (164) reports that the induction of a prolonged anger state by teasing increases the likelihood of a rat's convulsing upon exposure to loud sound. In the same laboratory, Arnold (2) observed that on trials in which "fear" reactions (tremor, twitching, lip-wetting) occur there is decreased probability of seizure. That injections of adrenin reduce susceptibility is presumably related to this finding; the adrenin induces the fear symptoms, which becomes a substitute for convulsion.

Other Behavioral Relationships

Substitute activities. If under "convulsion" is subsumed only the running-tonic-clonic pattern, then an inverse relation obtains between the appearance of convulsions and certain other forms of response to sound. In nonsusceptible rats, or on the nonseizure trials of susceptible rats, the onset of the auditory stimulus is frequently followed by face rubbing, ear scratching, body cleaning, sniffing, or chattering (2, 56, 88, 97, 151). Inasmuch as seizure virtually never ensues if such forms of behavior occur, they may be classified as substitute forms of response—in one sense, a means of protection against convulsion.

Other forms of alternate activity have been described as characterizing non-convulsers: "haphazard" movements (4), increased motor activity such as climbing (10), exploration (2), attack of the sound source (58), general "nervousness" (87, 92). Bitterman and Warden (15) point out that adding mild shock to sound stimulation not only produces more activity in the rat than results from sound stimulation alone, but also is followed by fewer convulsive attacks. It has been suggested that such forms of behavior, as well as chronic emotionality (11), fear reactions (2, 3), and forced exercise (76, 101), drain off the excitation provoked by intense stimulation before it can build up to an explosive level (161). Yet to be discovered are the factors which predispose an animal to these responses rather than to seizure,

Escape reactions. When the rat is forced (e.g., by an air blast) to make difficult discriminations in the Lashley apparatus, the well developed response is often disrupted by the appearance of position habits and inaccurate ("abortive") jumping. In the former, no attempt is made to select the correct stimulus card; instead, the jump is always to the same side. In the abortive reaction, the animal gradually loses his skill in reaching the window, hitting above, below, to one side, or even falling short of the food platform. These responses appear particularly when the negative stimulus is presented, and ordinarily can be eliminated by retraining. Maier (93) and Maier and Klee (103) report that the number of seizures produced by conflict plus sound diminishes when an animal develops either of these modes of jumping. The explanation suggested is that these reactions offer a means of escape from

the problem situation. An animal for whom the situation no longer involves choice (fixated position habit) or for whom punishment for incorrect choice is decreased (abortive jumping) experiences less intense conflict, and so, granting the facilitating role of conflict or emotional upset, would be expected to exhibit fewer convulsions.

The possibility of escape of another sort was provided in a sound-stimulation experiment lacking the discrimination problem (56). In this situation, the rat was exposed to the loud sound of a bell and during the running phase of the seizure was given the opportunity to run into a small "rat-hole." It is reported that of the rats entering this shelter, fewer continued into the convulsive stage of the seizure than had done so in trials prior to the introduction of this feature. The results are interpreted to mean that the rat's "affective tone" is a factor in determining convulsive behavior, and that the "psychological shelter" response makes the badly frightened rat feel more secure. Evaluation of the observations is difficult without further information regarding the control of such factors as sound intensity and adaptation.

Activity and restraint. Numerous observations have been made which relate the probability of sound-induced seizure to the amount of gross movement allowed or required during or just preceding the stimulation. If the animal is forced to move vigorously during this time, the convulsive responses decrease in number (76, 101) or are inhibited completely (94). On the other hand, placing the rat in a relatively small cage, so that movements tend to be restricted, is reported to increase the percentage of seizures (99, 101), although the meaning of these data is somewhat obscure because of small numbers and limited controls. This trend apparently breaks down upon increasing the degree of restraint still further, for binding the subject during auditory stimulation reduces (72, 111) or eliminates (56) the appearance of seizures, as well as certain physiological manifestations of abnormal response (88). It may be that while the relative restraint of a small cage reduces neuromuscular activity, the restraint of close binding actually increases neuromuscular activity in the form of isometric contractions against the bonds, thus providing as much of a "safety valve" as does forced exercise or substitute behavior. It would be enlightening to determine if habituation to such restraint leads to the reappearance of convulsive symptoms.

Nutritional Factors

The precise influence of nutritional factors upon convulsions induced by electroshock, sound stimulation, or conflict is difficult to determine by the average psychologist because of his lack of technical knowledge concerning the laboratory diet. It is often assumed that Purina chow, perhaps supplemented occasionally by a ration of spinach, is adequate for maintaining the rat at his physiological optimum, and

that for this reason the dietary factor can be disregarded as a variable relevant to convulsive behavior. There is accumulating, however, a formidable array of data on the effect of either extreme of dietary completeness: convulsions are very readily induced when certain constituents are minimal, and become increasingly difficult to provoke with the provision of generous supplements. As the influence of the extremes becomes more obvious, one is led to suspect that in the intermediate condition—that of “standard” colony feeding—the nutritional state is an omnipresent determiner. It would appear that more detailed examination of this factor is justified, extending even beyond studies of convulsions to such problems as learning and motivation.

Magnesium. Some fifteen years ago it was reported (85, 124) that rats on a low-magnesium diet frequently exhibited convulsions outwardly similar to the seizures later described as “neurotic” or “audio-genic.” Further experiments (50, 51, 52, 53, 135, 136, 165) have amplified our knowledge of this phenomenon. If a group of rats, free of convulsive responses even to intense sound, are put on a diet which includes only 1-2 mg. of magnesium per 100 g. of dry food, the plasma and body magnesium levels drop, vasodilatation is observed for a few days, and then a state of hyperexcitability begins. In this condition, all rats will exhibit the running-tonic-clonic pattern when stimulated by sound (but not when such stimuli as shock or light are used). At this stage the replacement of magnesium in the diet alleviates the symptoms in most of the animals (136). When smaller amounts are included in the diet (e.g., 0.4 mg. per 100 g. of food), the convulsive behavior may be induced sooner, is somewhat more severe in form, eventually comes to occur in the absence of any demonstrable stimulus, cannot be cured by the addition of magnesium salts, and often terminates in death (53). With 5 mg. or more of magnesium per 100 g. of food, some rats exhibit the convulsions only under special conditions, e.g., when lactating or when calcium is added to the diet. Supplements of vitamin G delay the appearance of the syndrome (53).

Vitamin B. In 1938, Kline, Tolle and Nelson (84), employing the rat as an assay object for vitamin B₁, described a polyneuritis in rats maintained for 25-50 days on a diet deficient in this vitamin. In the early stages “a state of extreme contraction of the musculature” was observed; later, “the animal . . . [exhibited] almost continuous uncontrolled movements” (p. 309). A 2-unit administration of the vitamin caused the symptoms to disappear within 24-48 hours. More closely resembling the sound-induced convulsive pattern was that observed in 1940 by Chick, El Sadr and Worden (17) in rats deprived of vitamin B₆. The seizures often occurred daily, and “the slightest external stimulus seemed to precipitate them” (p. 597). Cure was effected by the addition of B₆ in adequate amounts. A number of later studies (19, 24,

59, 127, 132, 133) have verified that the lack of normal B-vitamin content in the diet brings about susceptibility to sound-induced convulsions, and that in extreme cases "spontaneous"¹⁰ seizures occur that are indistinguishable in form from the audiogenic seizure. Conversely, almost complete protection against audiogenic seizure is afforded by the inclusion in the diet of large amounts of vitamin B, although the supplement is maximally effective only if provided early in life.

There are a number of complications that make difficult an exact statement of the role of the vitamin B complex in this behavior. Hunger (38, 121) and inanition (81, 132) themselves affect susceptibility to sound-induced seizure (each function passing through a maximum), and a subnormal vitamin level usually decreases food intake. Further, the effects of inanition and vitamin deficiency may cancel each other out when they are present simultaneously (131). Multiple vitamin deficiencies may similarly counteract each other, and supplementary feeding with other vitamins at times increases the severity of attack in B₆-deficient animals (134). Finally, restoration of food after a period of fasting leads first to an increase and then to a decrease of seizures (132). But certainly it appears that the dietary factor is an extremely important one in the determination of convulsions, and may account for the differences in susceptibility accompanying differences in age, strain, and colony background. The importance of this variable in unsuspected cases is enhanced by the observation that a deprivation which has no effect upon adult females may still induce susceptibility in their suckling young (24, 128, 134).

Other constituents. Other nutritional factors have been investigated for their effect upon response to auditory stimulation. Rats maintained on a vitamin D-deficient diet for the first four weeks of life were no more susceptible to sound than were a group given the standard laboratory diet (153). Increments in frequency of convulsive attack may result from parathyroidectomy and low calcium intake (50).

Although the mechanisms by which nutritional factors modify convulsive susceptibility are not yet known, it is clear that their influence upon the rat's physiological condition must constantly be

¹⁰ Whether or not an auditory stimulus is essential as a precipitant of convulsions in the malnourished rat is not clear. Patton, Karn and King (133) observed that the onset of seizures is often delayed until 1-5 min. after termination of the auditory stimulus, which may account for the "spontaneity" of some cases. Only the deafening of rats exhibiting allegedly spontaneous seizures can rule out this type of etiology, or strengthen the unique position of the auditory modality in the production of seizures. It may be relevant to note that Galambos and Davis (44) found about half the fibers in the cat auditory nerve firing "spontaneously" even when precautions had been taken to eliminate ambient noise. Incidentally, the overt manifestation of seizures in the discrimination situation has occasionally been observed to have its onset several seconds after the termination of the auditory stimulus (93).

considered by the investigator of seizure behavior. The manipulation of this variable gives promise of yielding results of broad significance.

Drugs

Allied to the modification of physiological condition through control of diet is the injection of drugs as a means of altering seizure susceptibility. There are a number of drugs, notably metrazol and insulin, which have been used as the precipitant of convulsions in the rat (68, 80, 141, 142, 174; cf. also Section II). The pattern of this response differs from that involving intense sound or low intensity electric shock chiefly in its lack of a running phase and, with metrazol, the prominence of its characteristic spinal torsion.

Metrazol, when given in subconvulsive doses, so affects the rat that an auditory stimulus otherwise ineffective becomes adequate to evoke a prompt seizure (56, 78, 106, 145, 154). Such an attack includes a running component, and in this way resembles the usual sound-induced pattern more closely than it does the metrazol convulsion, although its severity is often greater (78). The tonic flexion typical of the drug convulsion is observed only occasionally with this combination (106). Other drugs successfully utilized to lower resistance to sound include strychnine (2, 153, 154), caffeine sodium benzoate (154), sodium phosphate (50), coramine (166), eserine in adequate dosage (72, 73), nicotine (72, 74), and perhaps sodium bromide administered prenatally (64).

A few substances reduce susceptibility to sound-induced seizures: adrenin (2, 87), dilantin sodium (21, 54, 148), adrenal cortex (54), atropine (72, 73), and small doses of eserine (74). Rats rendered susceptible to sound stimulation by strychnine injections are protected from convulsion by adrenin (1), but the simultaneous use of atropine and eserine does not yield a similar neutralization (74). Semiparalyzing dosage of curare appears to abolish the susceptibility of rats on normal laboratory rations (88), but those maintained on a low-magnesium diet can still exhibit seizures under the influence of subparalytic amounts of curare (53), as do thyro-parathyroidectomized rats on a low calcium diet (50). Phenobarbital (77), mecholyl, benzyl benzoate, and pituitrin (74) fail to affect significantly the response to auditory stimulation.

Convulsive threshold to cerebral electroshock is uninfluenced by adrenin (125), atropine, calcium gluconate, thromboplastic suspension of brain substance, and vitamin K (66). The susceptibility is perhaps lowered slightly by 1.0 cc. injections of alcohol (125), and light ether anesthetic eliminates the convulsions (66). Animals rendered sensitive to sound by magnesium deprivation are protected temporarily by sodium amytal, while the midbrain drug picrotoxin perhaps enhances the probability of seizure (53).

In outlining the convulsant properties of drugs, it is always essential

to add that the "typical" effect of a particular drug can be altered by a great number of characteristics of the organism and of the external situation (69). These additional factors must be taken into account by the investigator adopting the pharmacological approach to convulsive behavior.

Previous Auditory Stimulation and Convulsion

It is seldom possible to elicit the full convulsive pattern by renewed stimulation directly following a seizure induced by auditory excitation. Certain exceptions to this trend appear, and in some cases of severe dietary deficiency a series of repetitive convulsions resembling status epilepticus may be observed. (As has already been pointed out, no such refractory period follows the electroshock convulsion.) In general, however, it appears that immediate susceptibility is greatly reduced by convulsion, and even by auditory stimulation unaccompanied by seizure (101). But the various investigations of this adaptation process as it extends over longer periods of time, involving repeated trials, yield divergent results. Several experimenters (11, 64, 76, 153) find that with repeated stimulation, usually at 24-hour intervals, the likelihood of the rat's response becomes progressively less, and that a vacation from stimulation restores the original level of response. On the other hand, Patton and Karn (130) and Parker (126) found little evidence of adaptation, even when the daily trials were continued for 30 days, and Griffiths (56) found that repeated stimulation actually made the animals less resistant to convulsion, as indicated by a shortened latency and increased duration of attack. Using the air blast as stimulus, Maier and Glaser (99) found a gradual decrease in number of attacks, but with jingling keys as the source of stimulation there was little decrement in frequency. These seemingly contradictory findings may be reconciled if it is assumed that a rat possesses a threshold of susceptibility which is a function of a number of factors and which is raised temporarily as the result of convulsion or stimulation. If adequate time elapses before the next trial to permit recovery of the threshold to its original level, a second response can occur. If, on the other hand, the second test is given before the threshold has returned to its former level, the convulsive response may fail to appear. This presumption is confirmed by the observation that animals regularly responding at 24-hour intervals fail to respond when the trials are given at shorter intervals (38). Of course, even when the recovery is incomplete, a more intense stimulus may nevertheless induce seizure. It is this possibility that explains why in one testing situation adaptation takes place while in another it is not evident; the rats in Patton and Karn's and Parker's experiments may have been stimulated more intensely than were those animals demonstrating adaptation, and Maier and Glaser's key jingling may have provided more effective (high frequency) stimulation than did their air

blast. In summary, the probability of a rat's response depends not only upon such factors as his genetic constitution, age, emotional state, and dietary background, but also upon the interval since the last testing, the intensity of the auditory stimulus, and his individual rate of recovery. Why it is that threshold should vary with experience is a challenging problem for the neurophysiologist.

Neural Ablation

One of the most obvious techniques by which the nature of convulsions can be studied is operation on the nervous system. The role of the auditory receptors, the participation of the cerebral cortex, the influence of lower brain centers, autonomic involvement—these are theoretically open to surgical investigation. Unfortunately, technical difficulties exist which have tempered the enthusiastic adoption of this approach, and as a consequence little information is available.

Cerebral lesion. Maier's original report (92) included three cases in which convulsive responses to difficult discrimination plus air blast were observed in spite of partial cortical extirpation (15–20% of the frontal cortex in two rats, and 25% of the median region in the third). According to Smith (153), "certain results suggest that removal of the frontal pole of the cortex abolishes the convulsive seizures" (p. 326). Two experiments have attacked the problem directly with conflicting results. Beach and Weaver (7) subjected the three most susceptible of a group of 12 rats to two operations during which was removed more than 90% of the neopallium, as well as most of the corpus callosum, much of the hippocampus, and parts of the corpus striatum. When tested about two weeks later, the decorticates responded on each of 12 tests to jingling keys, as contrasted with 91% of 42 preoperative tests. Unfortunately, the control animals had also increased in susceptibility during the interval, responding 79% of the time as compared to the previous 36% positive trials. In each group the proportion of seizures which included the tonic-clonic phase increased. Weiner and Morgan (169), on the other hand, found that while the percentage of seizures in their surgical control group did not change significantly, there was a decrease in the number of seizures exhibited by the 20 experimental animals. Among those suffering prefrontal lobectomy, the number of responses dropped from 22 to 10; in those with lesions in the auditory area, there was a reduction from 41 to 4; those with motor area lesions had 10 as contrasted with a previous 30 attacks. It is significant, however, that only one seizure in the prefrontal group included the tonic-clonic component of the convulsion, while all seizures appearing in the other operated groups were of the full convulsive type. No change in average latency of seizure was found, in contrast to the marked decrease reported by Beach and Weaver.

To reconcile these divergent results seems impossible. The differences in number and strain of rats, original level of susceptibility, operative recovery period, and extent of destruction qualify any direct comparison of the two experiments. Further definitive work will be necessary.

Vagotomy. The possible mediation of convulsions by the autonomic nervous system suggests operative limitation of its function. It seems on the basis of

preliminary data that vagotomy decreases (72) or abolishes (88) the appearance of seizures.

Other Factors

There are a number of minor factors which apparently contribute to individual differences in susceptibility. Some are subject to experimental manipulation, and others can be controlled only statistically. A full explanation of convulsive behavior must take them into account.

Blood composition. The relationship between plasma magnesium level and convulsion has already been noted (51, 165). Lacey (86) found that the serum protein level of rats responding convulsively to the buzzer and air blast was higher (significant at the 4% level) than that of nonsusceptibles and suggested neural dehydration as the mediating mechanism (cf. 38). Further, the susceptible group was characterized by the greater variability in blood sugar, perhaps relating to a dysfunction of the sympathetico-adrenal system.

Endocrine action. Little work has been done relating the endocrine balance to seizure threshold. Removal of the thyroid and parathyroid glands, coupled with low calcium diet, tends to increase susceptibility to stimulation-convulsion (50). No effect of oestrous cycle was observed by Mirsky, Elgart and Aring (117), which suggests lack of pituitary-ovarian influence.

Heart rate. In a group of rats selected for their convulsive response to air blast, Lindsley, Finger and Henry (88) found that in trials culminating in convulsion the initial heart rate averaged 379.4 per minute, increasing 20.1 beats per minute after the onset of stimulation, while in nonseizure trials the initial rate was 398.2 and the average increase only 4.4. These intertrial differences were not statistically reliable, however.

Startle response and general activity. Turner (167) found no difference between susceptibles and nonsusceptibles with respect to diurnal activity pattern. The startle response of the susceptibles was clearly the greater, however, as was the amount of their activity during brief periods of moderate noise.

Learning ability. In an exploratory study of the relationship between learning ability and seizure susceptibility, Sisk (152) gave 35 rats 14 daily trials on a 15-unit T-maze and then tested their convulsive reaction to the sound of a bell. The six responding positively to the noise had exhibited an average maze performance slightly inferior to that of the nonsusceptibles, but the difference was too small to be more than suggestive.

Physical condition. Farris and Yeakel (34), finding that rats tested in the spring and summer reacted more violently to sound than did those tested in the winter, have suggested that the significant variable is general physical condition: the more susceptible rats were from litters born in late fall and winter, and typically less sturdy. (Cf. also 95, 131, and *Nutritional Factors* above).

Body temperature. There seems to be a tendency for the lowering of body temperature to decrease (97, 101) or abolish (94) seizures in response to sound, although the experimental demonstrations are somewhat complicated by other factors.

Combinations. In an attempt to establish physiological patterns distinguishing between rats responding convulsively to noise and nonsusceptibles, Marcuse

and Moore (110) compared measures of their retiring behavior, sensitization, weight, frequency of defecation, and cardiac action. By using a composite criterion including at least two of these indices, "... a group was selected which comprised 70% of the susceptibles and 18% of the non-susceptibles . . . " (p. 2).

V. THE EFFECT OF CONVULSIONS

Because of the possible applications to problems of human shock therapy and epilepsy, a number of observations have been made of the effect of convulsions upon other behavior patterns and the structure of the rat. Most of the specifically planned experiments have been concerned with learning and retention, but incidental observations are recorded regarding several other points.

Learning Ability

The few experiments in which the effect of convulsions upon learning has been investigated point generally to a slight decrement in acquisition ability, especially as measured by speed of response (16, 71, 141, 149). However, the differences are so slight, the scope of the research so limited, and the operation of distorting factors so probable (e.g., motivation, 6, and emotional state, 149), that this conclusion can be presented only very tentatively. The results obtained in future experiments will undoubtedly depend upon such variables as the kind of convulsing situation employed, the age and physiological condition of the rats, the number and spacing of the seizures, the difficulty of the learning problem, and the interval between the convulsion series and the learning test.

Retention

The disrupting effect of convulsion upon material already acquired has been subjected to somewhat closer scrutiny. Four investigators have employed loud sound to induce the convulsions. Three experiments (55, 151, 160) revealed no appreciable loss of a maze habit after 1-3 convulsions, although there was in some cases a tendency for the time of running to increase unless some form of forcing was adopted. In one experiment (108) rats that had had 1-5 (?) seizures in the food box of the maze made significantly more errors on relearning trials than did those animals stimulated similarly but having had no seizure. The experimenter, however, felt that the deterioration in maze performance "was a product of the superposition of fear on a learned pattern, rather than a loss of a learned pattern" (p. 34). Taken together, these results

suggest that although there may be a change in general behavior as the result of a limited number of sound-induced seizures, there need be no disorganization of recently learned maze habits. Of course, there is no reason to assume similar stability of response following a longer series of convulsions.

The deleterious effect upon retention of even one or two convulsions induced by metrazol or insulin seems quite definite. In only one investigation (68) did the convulsed animals respond as well upon retest as did the control group, and the significance of this single finding is limited by the small number of convulsions involved (average of two). In other investigations involving maze behavior the seizures did lead to deterioration of the habit, particularly so with more difficult, less well learned, and more recently acquired responses (9, 89, 141). Insulin and metrazol convulsions have also been shown capable of effecting the recovery of an extinguished conditioned response (46, 80). On the other hand, there is no indication that either a fixation in the Lashley jumping apparatus or the well-learned discrimination habit itself is affected by metrazol convulsion (82, 103).

The results of experiments employing electroshock are similarly positive under certain conditions. It appears that very simple habits (e.g., a running response, a two-choice T-maze) may not be impaired, at least as measured by the number of errors on the retention trial (149, 160). However, on more difficult mazes and in one discrimination situation, retention scores have been lower, and certainly so if the number of shocks is increased (29, 71, 156). In at least one instance the deterioration of the pattern seemed to be due not just to emotional factors, but to a "direct effect on cerebral tissue" (29, p. 278). In still another experiment, an extinguished conditioned response reappeared after a single electrically-induced convulsion (80).

As in the series of studies involving learning ability, these investigations of retention are too limited in range to be conclusive. It appears that, given enough convulsions, there is measurable interference with certain habits. An extension of these experiments is sorely needed, with variation of difficulty and age of the pattern, number and spacing of practice trials, number and type of convulsion, retention interval, and method of retesting. And as Stainbrook (158) has intimated, it is always essential to differentiate between actual loss of habit (perhaps indicated by the criterion of errors or number of relearning trials) and the generalized effect of emotion or lack of alertness (as indicated by time measures). The obvious corollary of these studies is the investigation of the neurophysiological changes paralleling the behavioral modifications.

Reasoning Ability

Sharp, Winder and Stone (146) studied the effect of electroshock upon "reasoning ability" as measured by Maier's apparatus (90) which requires "... bringing together spontaneously two elements of past experience without having them previously associated by contiguity" (91, p. 46). The tests were given $1\frac{1}{2}$ or 22 hours after electroshock convulsion. Although only six subjects were used and individual differences were apparent, the trend was in the direction of decreased distance in exploratory runs, increased time for test runs, and decreased accuracy scores. The impairment was greater $1\frac{1}{2}$ hours after shock than when measured after 22 hours, indicating that at least part of the disturbance was of a transitory nature, similar in this respect to the effect appearing in retention studies.

Emotional Patterns

In the course of experiments testing learning and retention following convulsion, some observations have been made of changes occurring in other behavior patterns, particularly those described as "emotional." Stainbrook (156, 160), for example, reports that after a series of electroshocks, an "experimental neurosis" appeared, during which the rats manifested excitement and resistance when approached, and appeared upset by slight noises. The condition persisted for several months. Page (125) found that while electrically convulsed animals resisted the experimenter during the first few days, they soon became passive and lethargic. In two experiments, one employing shock (149) and the other air blast (108), the rats seemed trying to avoid the situation in which the convulsions had been experienced, and evidenced emotional disturbance when placed in the situation again. Directly contradictory is the observation of Duncan (29) that animals having undergone 30 electroshock convulsions evinced no fear in the experimental situation, in marked contrast to rats shocked on the legs. Most experimenters employing loud sound as the convulsive agent similarly report no avoidance of the situation by the rat (e.g., 92).

This survey of the available data makes it clear that too little is known about the generalized behavioral changes following convulsion in the rat. It is desirable that future experiments be oriented more deliberately and systematically in this direction, with objective measures adopted to reveal the possible changes.

Activity

No very refined technique is necessary to detect the decrease in activity immediately following convulsion. This relative immobility

varies in degree and duration from seizure to seizure, often retarding maze running for a period of several hours. Only three attempts at quantitative evaluation of activity changes have been reported. In the one involving sound-induced seizures (42) it was found that auditory stimulation without convulsion did not affect subsequent activity, but following a seizure the activity level dropped significantly. Recovery was apparently complete within ten hours, with no cumulative effect discernible after three seizures. Unfortunately, the number of attacks induced was too few to support any generalization regarding long-range modification of diurnal pattern.

It appears that a single electroshock likewise depresses activity measurably, and that during a period of daily convulsions the activity level drops significantly below normal (162, 170). The effect is not long-lasting, however; some recovery is evident in the first few days following termination of the series, and the preconvulsion level is approached within 2-4 weeks.

Mating and Maternal Behavior

Some of the early investigators reported difficulty in breeding rats shown to be susceptible to sound-induced convulsion (e.g., 100). This might seem to imply that the convulsions themselves exert a depressing influence upon mating. The most direct test of this supposition is an experiment by Farris and Yeakel (32, 35). They found that animals subjected to air blasting five days a week beginning at weaning produced 30% more litters than did control litter mates not subjected to the sound. On the other hand, significantly fewer offspring of these experimental animals than of the controls survived the first day of life, due in part to a noticeable inefficiency of maternal behavior among the experimental mothers (cf. 57). Unfortunately, no breakdown of figures is given that would permit a separation of the influence of the auditory stimulation itself from that of the convulsions produced thereby.

Physiological Changes; Death

To the extent that behavior changes are found to result from the experiencing of convulsions, parallel changes in the structure or physiology of the organism might be expected. In general, no very striking alteration has been found, perhaps because relatively little attention has been paid the problem.

Rats (as well as other laboratory animals) undergoing repeated convulsions, induced either electrically or by loud sound, gradually deteriorate in general physical condition and gain less weight than do

control animals (42, 125). It is possible that the systolic blood pressure becomes abnormally high (115). According to one exploratory study, exposure to sound, with or without seizure, causes the deposition of pigment granules in the adherent mucus of the stomach walls, but with no evidence of ulceration (104). Electroshock convulsion (and not the shock itself) produces hemorrhages in the pia-arachnoid and the parenchyma, presumably because of circulatory changes accompanying the convulsion (66). Finally, studies of the blood indicate that, following audiogenic seizure, "the gray Norway rat rapidly developed a shorter blood coagulation time, and an acidosis caused by acids other than carbonic in the blood" (22, p. 525). It will be noted that most, if not all of these physiological changes are of a temporary nature. Direct evidence of permanent modifications, e.g., in the central nervous system, is lacking.

The probability of death in convulsion apparently depends for the most part upon the age and physiological condition of the subject. Ordinarily, very few deaths occur during sound-induced seizure, although up to 4% mortality has been recorded among very young animals (4, 153). In convulsions associated with nutritional deficiency, death seldom occurs if the deficit is moderate (17), but in extreme cases of magnesium deprivation death during convulsion, within 30 days, is the virtually invariable outcome (53, 85).

VI. PHYSIOLOGICAL BASES OF CONVULSIONS

When the psychologist is puzzled by the appearance of an atypical pattern of behavior, he often looks to the physiology laboratory for an answer. Or at least he remains uneasy until an answer to his "why" can be phrased in physiological language. An imposing array of behavioral observations confronts the student of rat convulsions, but the interrelationships of the numerous variables are so complex (and so frequently ignored experimentally) that a coherent picture is still lacking. Unable to systematize his behavioral findings at the gross level, the investigator searches for a description at the neurophysiological level, to which the diverse external manifestations may be referred. In short, the first five sections might be considered a mere preliminary to this section, which should be the climax of the review.

Unfortunately, however, no thoroughgoing theory of convulsion has been attempted, nor shall I essay one here. In reality, the prime purpose of this report has been to present the problem in such a way that the reader with an adequate background will undertake his own formulation. It may be of value to mention again some findings that

seem particularly deserving of attention in such a theory, and to add a few conjectures that have been made regarding them. But all this is intended only to hint at one direction which speculation may take, and is certainly too tentative to be a restraining influence on the flow of hypothesis. Even the contradiction of experimental data need be no final deterrent, for a great deal of the work demands repetition with more satisfactory controls.

It would appear that in the convulsions attributed to conflict, auditory stimulation, minimal electroshock, and nutritional deficit, the neural discharge finds a final path that is approximately common. If only the tonic-clonic stage were considered, this statement could also include certain drug-induced seizures. That this intense discharge is central in origin is suggested by several lines of evidence.

1. Electroencephalographic recording before and during audiogenic seizure reveals cortical involvement (88).
2. Cortical extirpation may decrease the probability of seizure to sound (169).
3. The mode of interaction of the convulsion-facilitating eserine and the inhibitory atropine indicates that their influence is through the central nervous system (73, 74).
4. Coramine, which increases the reactivity of the central nervous system, in subconvulsive doses increases convulsive susceptibility to sound (166).

But there is also reason to believe that subcortical and autonomic centers contribute to the innervation.

1. Cortical extirpation does not always abolish seizures (7).
2. Although the total number of seizures may be diminished by lesions in the motor cortex, the proportion of attacks which include the tonic-clonic phase is thereby increased. Weiner and Morgan (169) suggest that the normal function of this region tends to inhibit the more incoordinated (convulsive) discharge which has a subcortical origin.
3. According to Morgan (119) and Weiner and Morgan (169) latency of seizure is uninfluenced by a number of factors, including cortical extirpation, which affect threshold of seizure. This timing process may thus be a subcortical function. Relevant to this point is Smith's suggestion (153) of a physiological timing mechanism, cyclic in character.
4. Changes in heart rate during auditory stimulation are prognostic of susceptibility during that trial. Both these changes and the susceptibility to seizure are minimized by vagotomy (88).
5. In metrazol and electroshock convulsion the autonomic system is found to be stimulated, with the action of the sympathetico-adrenal system predominating (37, 80).

The puzzle is: Why do the central overloading and the convulsive synchrony of central firing occur? A number of possibilities may be mentioned.

1. Relative cerebral anoxia seems to accompany convulsion induced by electroshock and certain drugs (25, 79). Such a condition may well alter the cell membrane surfaces so that the tissue becomes abnormally sensitive to stimulation.
2. In states of malnutrition a parallel sensitization might take place, so that normally ineffective stimuli would serve to cause mass response (cf. 136).
3. In instances where these factors are usually absent, auditory stimulation might induce a temporary anoxia. Whether the mechanism involved should be termed reflex or emotional cannot be stated now.
4. Due to the proximity of the auditory and the motor centers, direct spread of excitation may be uniquely facilitated when this modality of stimulation is employed.
5. Hypothalamic centers may be excited rather directly, discharging both centrally and peripherally, via the autonomic nervous system (88). In the intact organism reverberation from the visceral organs may supplement the original cortical innervation, leading to explosive overflow unless constantly drained off by emotional (2, 114) or substitute (56, 76, 88) reactions.

It is obvious that these suggestions are little more than isolated phrases in the physiological interpretation that finally must be made. More work, particularly involving electrical, pharmacological, and operative techniques, must be planned with neurophysiological hypotheses as the guiding orientation, at the same time recognizing the multiplicity of factors that must be segregated and controlled. It appears that breaking down the classificatory distinctions between at least some of the "types" of convulsion would facilitate this approach. What the final categories will be, and their relation to human paroxysmal behavior, must wait on a more complete experimental integration.

BIBLIOGRAPHY

1. ARNOLD, MAGDA B. Emotional factors in experimental neuroses. *Bull. Canad. psychol. Ass.*, 1942, 2, 25.
2. ARNOLD, MAGDA, B. Emotional factors in experimental neuroses. *J. exp. Psychol.*, 1944, 34, 257-281.
3. ARNOLD, MAGDA B. Physiological differentiation of emotional states. *Psychol. Rev.*, 1945, 52, 35-48.
4. AUER, E. T., & SMITH, K. U. Characteristics of epileptoid convulsive reactions produced in rats by auditory stimulation. *J. comp. Psychol.*, 1940, 30, 255-259.
5. BARD, P. Emotion: I. The neuro-humoral basis of emotional reactions. In C. Murchison (Ed.), *Handbook of general experimental psychology*. Worcester: Clark Univ. Press, 1934. Pp. 264-311.
6. BAYROFF, A. G. Air blasts as substitutes for electric shock in discrimination learning of white rats. *J. comp. Psychol.*, 1940, 29, 109-118.
7. BEACH, F. A., & WEAVER, THELMA. Noise-induced seizures in the rat and their modification by cerebral injury. *J. comp. Neurol.*, 1943, 79, 379-392.
8. BEAUVILLAIN, A., & BOILOT, YVONNE. Action convulsivante de l'insuline

- sur le rat albinos. Application au titrage de l'hormone. *Compt. rend. Soc. Biol.*, 1942, **136**, 322.
9. BERMAN, L., & RIESS, B. F. The effect of insulin shock on learning in the white rat. *Science*, 1942, **95**, 511-512.
 10. BERNHARDT, K. S., TOBIN, F. J., & SIGNORI, E. Exploratory studies of abnormal behavior in the rat. *J. comp. Psychol.*, 1941, **32**, 575-582.
 11. BILLINGSLEA, F. Y. The relationship between emotionality and various other salients of behavior in the rat. *J. comp. Psychol.*, 1941, **31**, 69-77.
 12. BITTERMAN, M. E. Behavior disorder as a function of the relative strength of antagonistic response tendencies. *Psychol. Rev.*, 1944, **51**, 375-378.
 13. BITTERMAN, M. E. A reply to Dr. Finger. *Psychol. Rev.*, 1946, **53**, 116-118.
 14. BITTERMAN, M. E. Sound-induced seizures in the rat: psychological factors in the etiology of the seizures. Unpublished. (Symposium, American Psychological Association meeting, 1946.)
 15. BITTERMAN, M. E., & WARDEN, C. J. The inhibition of convulsive seizures in the white rat by the use of electric shock. *J. comp. Psychol.*, 1943, **35**, 133-137.
 16. BUNCH, M. E., & MUELLER, C. G. The influence of metrazol upon maze learning ability. *J. comp. Psychol.*, 1941, **32**, 569-574.
 17. CHICK, HARRIETTE, EL SADR, M. M., & WORDEN, A. N. Occurrence of fits of an epileptiform nature in rats maintained for long periods on a diet deprived of vitamin B₆. *Biochem. J.*, 1940, **34**, 595-600.
 18. CHICK, HARRIETTE, MACRAE, T. F., MARTIN, A. J. P., & MARTIN, C. J. The water-soluble B-vitamins other than aneurin (vitamin B₁), ribo-
flavin and nicotinic acid required by the pig. *Biochem. J.*, 1938, **32**, 2207-2224.
 19. CHICK, HARRIETTE, MACRAE, T. F., & WORDEN, A. N. Relation of skin lesions in the rat to deficiency in the diet of different B₂-vitamins. *Biochem. J.*, 1940, **34**, 580-594.
 20. COBB, S. Causes of epilepsy. *Arch. Neurol. Psychiat.*, Chicago, 1932, **27**, 1245-1256.
 21. COHEN, L. M., & KARN, H. W. The anti-convulsant action of dilantin on sound-induced seizures in the rat. *J. comp. Psychol.*, 1943, **35**, 307-310.
 22. COLE, W. H., YEAKEL, ELEANOR, H., & FARRIS, E. J. A preliminary study of changes in the blood of gray Norway rats following audiogenic seizures. *Anat. Rec.*, 1942, **84**, 524-525.
 23. CRITCHLEY, M. Musicogenic epilepsy. *Brain*, 1937, **60**, 13-27.
 24. DANIEL, ESTHER P., KLINE, O. L., & TOLLE, C. D. A convulsive syndrome in young rats associated with pyridoxine deficiency. *J. Nutr.*, 1942, **23**, 205-216.
 25. DAVIS, E. W., McCULLOCH, W. S., & ROSEMAN, E. Rapid changes in the O₂ tension of cerebral cortex during induced convulsions. *Amer. J. Psychiat.*, 1944, **100**, 825-829.
 26. DAY, H. G., KRUSE, H. D., & MCCOLLUM, E. V. Studies on magnesium deficiency in animals. VII. The effects of magnesium deprivation, with a superimposed calcium deficiency, on the animal body, as revealed by symptomatology and blood changes. *J. biol. Chem.*, 1935-36, **112**, 337-359.
 27. DICE, L. R. Inheritance of waltzing and of epilepsy in mice of the genus *Peromyscus*. *J. Mammal.*, 1935, **16**, 25-35.
 28. DONALDSON, H. H. *The rat: data and*

- reference tables. (2nd Ed.) Philadelphia: Wistar Inst., 1924.
29. DUNCAN, C. P. The effect of electroshock convulsions on the maze habit in the white rat. *J. exp. Psychol.*, 1945, **35**, 267-278.
30. ENGLAND, A. O. Factors influencing electrically induced convulsions in rats. *J. Psychol.*, 1942, **14**, 227-234.
31. FARRIS, E. J., & YEAKEL, ELEANOR H. The effect of age upon susceptibility to audiogenic seizures in albino rats. *J. comp. Psychol.*, 1942, **33**, 249-251.
32. FARRIS, E. J., & YEAKEL, ELEANOR H. Effects of a neurophysiological stimulus on the breeding of albino rats. *Anat. Rec.*, 1942, **84**, 454.
33. FARRIS, E. J., & YEAKEL, ELEANOR H. Sex and increasing age as factors in the frequency of audiogenic seizures in albino rats. *J. comp. Psychol.*, 1942, **34**, 75-78.
34. FARRIS, E. J., & YEAKEL, ELEANOR H. The susceptibility of albino and gray Norway rats to audiogenic seizures. *J. comp. Psychol.*, 1943, **35**, 73-80.
35. FARRIS, E. J., & YEAKEL, ELEANOR H. Breeding and rearing of young by albino rats subjected to auditory stimulation. *Anat. Rec.*, 1944, **89**, 325-330.
36. FARRIS, E. J., & YEAKEL, ELEANOR H. Emotional behavior of gray Norway and Wistar albino rats. *J. comp. Psychol.*, 1945, **38**, 109-118.
37. FELDMAN, J., CORTELL, R., & GELLHORN, E. On the vago-insulin and sympathetico-adrenal system and their mutual relationship under conditions of central excitation induced by anoxia and convulsant drugs. *Amer. J. Physiol.*, 1940, **131**, 281-289.
38. FINGER, F. W. Factors influencing audiogenic seizures in the rat: repeated stimulation and deprivation of food and drink. *Amer. J. Psychol.*, 1942, **55**, 68-76.
39. FINGER, F. W. Factors influencing audiogenic seizures in the rat: II. Heredity and age. *J. comp. Psychol.*, 1943, **35**, 227-232.
40. FINGER, F. W. Experimental behavior disorders in the rat. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald Press Co., 1944. Pp. 413-430.
41. FINGER, F. W. Abnormal animal behavior and conflict. *Psychol. Rev.*, 1945, **52**, 230-233.
42. FINGER, F. W., & SCHLOSBERG, H. The effect of audiogenic seizures on general activity of the white rat. *Amer. J. Psychol.*, 1941, **54**, 518-527.
43. FORSTER, F. M. Action of acetylcholine on motor cortex; correlation of effects of acetylcholine and epilepsy. *Arch. Neurol. Psychiat., Chicago*, 1945, **54**, 391-394.
44. GALAMBOS, R., & DAVIS, H. The response of single auditory-nerve fibers to acoustic stimulation. *J. Neurophysiol.*, 1943, **6**, 39-58.
45. GALAMBOS, R., & MORGAN, C. T. The production of audiogenic seizures by interrupted tones. *J. exp. Psychol.*, 1943, **32**, 435-442.
46. GELLHORN, E., & MINATOYA, H. The effect of insulin hypoglycemia on conditioned reflexes. *J. Neurophysiol.*, 1943, **6**, 161-171.
47. GLASER, N. M. Autonomic changes associated with abnormal behavior in the rat. I. Analysis of changes in heart rate occurring as a result of responses in an auditory situation. II. The effect of metrazol upon heart rate. Unpublished Doctor's thesis, Univ. Mich., 1941.
48. GOLUB, L. M., & MORGAN, C. T. Patterns of electrogenic seizures in rats: their relation to stimulus-intensity and to audiogenic sei-

- zures. *J. comp. Psychol.*, 1945, **38**, 239-245.
49. GOULD, J., & MORGAN, C. T. Hearing in the rat at high frequencies. *Science*, 1941, **94**, 168.
 50. GREENBERG, D. M., BOELTER, MURIEL D. D., & KNOPF, B. W. Factors concerned in the development of tetany by the rat. *Amer. J. Physiol.*, 1942, **137**, 459-467.
 51. GREENBERG, D. M., & TUFTS, ELMA V. The effect of a diet low in magnesium on the rat. *J. biol. Chem.*, 1935, **109**, xxxviii-xxxix of Proceedings.
 52. GREENBERG, D. M., & TUFTS, ELMA V. Effect of a diet low in magnesium on the rat. *Proc. Soc. exp. Biol.*, N. Y., 1935, **32**, 674-675.
 53. GREENBERG, D. M., & TUFTS, ELMA V. The nature of magnesium tetany. *Amer. J. Physiol.*, 1938, **121**, 416-423.
 54. GRIFFITHS, W. J. The effects of dilantin on convulsive seizures in the white rat. *J. comp. Psychol.*, 1942, **33**, 291-296.
 55. GRIFFITHS, W. J. The persistence of convulsions in the white rat. *J. comp. Psychol.*, 1942, **34**, 279-283.
 56. GRIFFITHS, W. J. The production of convulsions in the white rat. *Comp. Psychol. Monogr.*, 1942, **17**, No. 8. Pp. 29.
 57. GRIFFITHS, W. J. Transmission of convulsions in the white rat. *J. comp. Psychol.*, 1942, **34**, 263-277.
 58. GRIFFITHS, W. J. Absence of audiogenic seizures in wild Norway and Alexandrine rats. *Science*, 1944, **99**, 62-63.
 59. GRIFFITHS, W. J. The effect of thiamine hydrochloride on the incidence of audiogenic seizures among selectively bred albino rats. *J. comp. Psychol.*, 1945, **38**, 65-68.
 60. HALE, W. M., & MCKEE, A. P. The intracranial toxicity of influenza virus for mice. *Proc. Soc. exp. Biol.*, N. Y., 1945, **59**, 81-84.
 61. HALL, C. S. Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. *J. comp. Psychol.*, 1934, **18**, 385-403.
 62. HALL, C. S. The inheritance of emotionality. *Sigma Xi Quart.*, 1938, **26**, 17-27, 37.
 63. HALL, C. S., & MARTIN, R. F. A standard experimental situation for the study of abnormal behavior in the rat. *J. Psychol.*, 1940, **10**, 207-210.
 64. HAMILTON, H. C. The effect of the administration of sodium bromide on the behavior of the offspring: IV. Emotionality (timidity) and experimentally induced seizures. *J. Psychol.*, 1945, **19**, 17-30.
 65. HAMILTON, JANE R. Epileptiform convulsions in rats. I. Description of the phenomena and a comparison with symptomatology of human epilepsy. *J. comp. Psychol.*, 1942, **33**, 297-303.
 66. HEILBRUNN, G. Prevention of hemorrhages in the brain in experimental electric shock. *Arch. Neurol. Psychiat.*, Chicago, 1943, **50**, 450-455.
 67. HENLE, GERTRUDE, & HENLE, W. Neurological signs in mice following intracerebral inoculation of influenza viruses. *Science*, 1944, **100**, 410-411.
 68. HERON, W. T., & CARLSON, W. S. The effects of metrazol shock on retention of the maze habit. *J. comp. Psychol.*, 1941, **32**, 307-309.
 69. HOLCK, H. G. O. Dosage of drugs for rats. In J. Q. Griffith & E. J. Farris (Eds.), *The rat in laboratory investigation*. Philadelphia: Lippincott, 1942. Pp. 297-350.
 70. HOOBLER, S. W., KRUSE, H. D., & MCCOLLUM, E. V. Studies on magnesium deficiency in animals. VIII.

- The effects of magnesium deprivation on the total and ultrafilterable calcium and magnesium of the serum. *Amer. J. Hyg.*, 1937, 25, 86-106.
71. HOROWITZ, M. W., & STONE, C. P. The ease of learning a new habit in relation to the disorganization of an interfering habit as affected by electro-convulsive shock in the rat. *Amer. Psychologist*, 1946, 1, 449.
 72. HUMPHREY, G. Experiments on the physiological mechanism of noise-induced seizures in the albino rat. *Bull. Canad. psychol. Ass.*, 1941, 1, 39-41.
 73. HUMPHREY, G. Experiments on the physiological mechanism of noise-induced seizures in the albino rat. I. The action of parasympathetic drugs. *J. comp. Psychol.*, 1942, 33, 315-323.
 74. HUMPHREY, G. Experiments on the physiological mechanism of noise-induced seizures in the albino rat. II. The site of action of the parasympathetic drugs. *J. comp. Psychol.*, 1942, 33, 325-342.
 75. HUMPHREY, G., & MARCUSE, F. New methods of obtaining neurotic behavior in rats. *Amer. J. Psychol.*, 1939, 52, 616-619.
 76. HUMPHREY, G., & MARCUSE, F. Factors influencing the susceptibility of albino rats to convulsive attacks under intense auditory stimulation. *J. comp. Psychol.*, 1941, 32, 285-306.
 77. JONES, M. R. Some observations on effects of phenobarbital on emotional responses and air-induced seizures. *J. comp. Psychol.*, 1944, 37, 159-163.
 78. KARN, H. W., LODOWSKI, C. H., & PATTON, R. A. The effect of metrazol on the susceptibility of rats to sound-induced seizures. *J. comp. Psychol.*, 1941, 32, 563-567.
 79. KESSLER, M., & GELLHORN, E. Effect of electrically induced convulsions on vago-insulin and sympathetico-adrenal systems. *Proc. Soc. exp. Biol., N. Y.*, 1941, 46, 64-66.
 80. KESSLER, M., & GELLHORN, E. The effect of electrically and chemically induced convulsions on conditioned reflexes. *Amer. J. Psychiat.*, 1943, 99, 687-691.
 81. KING, C. G., KARN, H. W., & PATTON, R. A. Nutritional deficiency as a factor in the abnormal behavior of experimental animals. *Science*, 1941, 94, 186-187.
 82. KLEE, J. B. The relation of frustration and motivation to the production of abnormal fixations in the rat. *Psychol. Monogr.*, 1944, 56, No. 4. Pp. iv+45.
 83. KLEIN, D. B. *Mental hygiene*. New York: Holt, 1944.
 84. KLINE, O. L., TOLLE, C. D., & NELSON, E. M. Vitamin B₁ assay by a rat-curative procedure. *J. Ass. off. agric. Chem., Wash.*, 1938, 21, 305-313.
 85. KRUSE, H. D., ORENT, ELSA, R., & MCCOLLUM, E. V. Studies on magnesium deficiency in animals. I. Symptomatology resulting from magnesium deprivation. *J. biol. Chem.*, 1932, 96, 519-539.
 86. LACEY, O. L. The dependence of behavior disorder in the rat upon blood composition. Audiogenic seizure as a function of blood composition. *J. comp. Psychol.*, 1945, 38, 257-270.
 87. LACEY, O. L. The dependence of behavior disorder in the rat upon blood composition. Behavior disorder as a function of repeated adrenalin injections. *J. comp. Psychol.*, 1945, 38, 277-284.
 88. LINDSLEY, D. B., FINGER, F. W., & HENRY, C. E. Some physiological aspects of audiogenic seizures in rats. *J. Neurophysiol.*, 1942, 5, 185-198.

89. LOKEN, R. D. Metrazol and maze behavior. *J. comp. Psychol.*, 1941, **32**, 11-16.
90. MAIER, N. R. F. Reasoning in white rats. *Comp. Psychol. Monogr.*, 1929, **6**, No. 3. Pp. 93.
91. MAIER, N. R. F. The effect of cerebral destruction on reasoning and learning in rats. *J. comp. Neurol.*, 1932, **54**, 45-75.
92. MAIER, N. R. F. *Studies of abnormal behavior in the rat: the neurotic pattern and an analysis of the situation which produces it*. New York: Harper, 1939.
93. MAIER, N. R. F. Studies of abnormal behavior in the rat: IV. Abortive behavior and its relation to the neurotic attack. *J. exp. Psychol.*, 1940, **27**, 369-393.
94. MAIER, N. R. F. Some factors which inhibit the abnormal reactions to auditory stimulation. *Psychol. Bull.*, 1942, **39**, 591.
95. MAIER, N. R. F. Studies of abnormal behavior in the rat. XIV. Strain differences in the inheritance of susceptibility to convulsions. *J. comp. Psychol.*, 1943, **35**, 327-335.
96. MAIER, N. R. F. Two types of behavior abnormality in the rat. *Bull. Menninger Clin.*, 1943, **7**, 141-147.
97. MAIER, N. R. F., & FELDMAN, R. S. Studies of abnormal behavior in the rat. XIV. Water spray as a means of inducing seizures. *J. comp. Psychol.*, 1946, **39**, 275-286.
98. MAIER, N. R. F., & GLASER, N. M. Experimentally produced neurotic behavior in the rat. Film. Bethlehem: A. Ford, 1938. 600 ft., 16 mm.
99. MAIER, N. R. F., & GLASER, N. M. Studies of abnormal behavior in the rat. II. A comparison of some convulsion-producing situations. *Comp. Psychol. Monogr.*, 1940, **16**, No. 1. Pp. 30.
100. MAIER, N. R. F., & GLASER, N. M. Studies of abnormal behavior in the rat. V. The inheritance of the "neurotic pattern." *J. comp. Psychol.*, 1940, **30**, 413-418.
101. MAIER, N. R. F., & GLASER, N. M. Studies of abnormal behavior in the rat. IX. Factors which influence the occurrence of seizures during auditory stimulation. *J. comp. Psychol.*, 1942, **34**, 11-21.
102. MAIER, N. R. F., & GLASER, N. M. Studies of abnormal behavior in the rat. X. The influences of age and sex on the susceptibility to seizures during auditory stimulation. *J. comp. Psychol.*, 1942, **34**, 23-28.
103. MAIER, N. R. F., & KLEE, J. B. Studies of abnormal behavior in the rat. VII. The permanent nature of abnormal fixations and their relation to convulsive tendencies. *J. exp. Psychol.*, 1941, **29**, 380-389.
104. MAIER, N. R. F., & PARKER, W. Studies of abnormal behavior in the rat. XVIII. Analysis of stomachs of rats repeatedly exposed to auditory stimulation. *J. comp. Psychol.*, 1945, **38**, 335-341.
105. MAIER, N. R. F., & SACKS, J. Studies of abnormal behavior in the rat. VI. Patterns of convulsive reactions to metrazol. *J. comp. Psychol.*, 1941, **32**, 489-502.
106. MAIER, N. R. F., SACKS, J., & GLASER, N. M. Studies of abnormal behavior in the rat. VIII. The influence of metrazol on seizures occurring during auditory stimulation. *J. comp. Psychol.*, 1941, **32**, 379-388.
107. MAIER, N. R. F., & WAPNER, S. Studies of abnormal behavior in the rat. XIII. The effect of punishment for seizures on seizure-frequency during auditory stimulation. *J. comp. Psychol.*, 1943, **35**, 247-248.
108. MAIER, N. R. F., & WAPNER, S. Studies of abnormal behavior in the rat. XV. The influence of maze behavior on seizures occurring during auditory stimulation and the effect

- of seizures on maze performance. *J. comp. Psychol.*, 1944, **37**, 23-34.
109. MARCUSE, F. L., & MOORE, A. U. Heart rate and respiration preceding and following audiogenic seizures in the white rat. *Proc. soc. exp. Biol., N. Y.*, 1941, **48**, 201-202.
110. MARCUSE, F. L., & MOORE, A. U. Experimental studies of physiological patterns in normal and abnormal animal behavior. *Bull. Canad. psychol. Ass.*, 1943, **3**, 1-3.
111. MARCUSE, F. L., & MOORE, A. U. Heart rate in the comatose state of audiogenic seizures. *J. exp. Psychol.*, 1943, **32**, 518-521.
112. MARCUSE, F. L., & MOORE, A. U. Tantrum behavior in the pig. *J. comp. Psychol.*, 1944, **37**, 235-241.
113. MARTIN, R. F. The incidence of Maier-type neuroses in emotional and nonemotional strains of rats. *Psychol. Bull.*, 1940, **37**, 582.
114. MARTIN, R. F., & HALL, C. S. Emotional behavior in the rat. V. The incidence of behavior derangements resulting from air-blast stimulation in emotional and non-emotional strains of rats. *J. comp. Psychol.*, 1941, **32**, 191-204.
115. MEDOFF, H. S., & BONGIOVANNI, A. M. Blood pressure in rats subjected to audiogenic stimulation. *Amer. J. Physiol.*, 1945, **143**, 300-305.
116. MERRITT, H. H., & PUTNAM, T. J. Sodium diphenyl hydantoinate in the treatment of convulsive disorders. *J. Amer. med. Ass.*, 1938, **111**, 1068-1073.
117. MIRSKY, I. A., ELGART, S., & ARING, C. D. Sonogenic convulsions in rats and mice. I. Control studies. *J. comp. Psychol.*, 1943, **35**, 249-253.
118. MORGAN, C. T. Review of N. R. F. Maier's *Studies of abnormal behavior in the rat*. *J. gen. Psychol.*, 1940, **23**, 227-233.
119. MORGAN, C. T. The latency of audiogenic seizures. *J. comp. Psychol.*, 1941, **32**, 267-284.
120. MORGAN, C. T., & GALAMBOS, R. Production of audiogenic seizures by tones of low frequency. *Amer. J. Psychol.*, 1942, **55**, 555-559.
121. MORGAN, C. T., & GOULD, J. Acoustical determinants of the "neurotic pattern" in rats. *Psychol. Rec.*, 1941, **4**, 258-268.
122. MORGAN, C. T., & MORGAN, JANE D. Auditory induction of an abnormal pattern of behavior in rats. *J. comp. Psychol.*, 1939, **27**, 505-508.
123. MORGAN, C. T., & WALDMAN, H. "Conflict" and audiogenic seizures. *J. comp. Psychol.*, 1941, **31**, 1-11.
124. ORENT, ELSA, R., KRUSE, H. D., & MCCOLLUM, E. V. Studies on magnesium deficiency in animals. II. Species variation in symptomatology of magnesium deprivation. *Amer. J. Physiol.*, 1932, **101**, 454-461.
125. PAGE, J. D. Studies in electrically induced convulsions in animals. *J. comp. Psychol.*, 1941, **31**, 181-194.
126. PARKER, M. M. Some fundamental characteristics of convulsions in rats. *Psychol. Bull.*, 1941, **38**, 579.
127. PATTON, R. A. The effect of vitamins on convulsive seizures in rats subjected to auditory stimulations. *J. comp. Psychol.*, 1941, **31**, 215-221.
128. PATTON, R. A. The effect of rice polish concentrate on the incidence of sound-induced convulsive seizures in young albino rats. *Amer. Psychologist*, 1946, **1**, 275.
129. PATTON, R. A. Sound-induced seizures in the rat: the study of constitutional and nutritional factors. Unpublished. (Symposium, American Psychological Association meeting, 1946.)
130. PATTON, R. A., & KARN, H. W. Abnormal behavior in rats subjected to repeated auditory stimulation. *J. comp. Psychol.*, 1941, **31**, 43-46.

131. PATTON, R. A., KARN, H. W., & KING, C. G. Studies on the nutritional basis of abnormal behavior in albino rats. I. The effect of vitamin B₁ and vitamin B-complex deficiency on convulsive seizures. *J. comp. Psychol.*, 1941, **32**, 543-550.
132. PATTON, R. A., KARN, H. W., & KING, C. G. Studies on the nutritional basis of abnormal behavior in albino rats. II. Further analysis of the effects of inanition and vitamin B₁ on convulsive seizures. *J. comp. Psychol.*, 1942, **33**, 253-258.
133. PATTON, R. A., KARN, H. W., & KING, C. G. Studies on the nutritional basis of abnormal behavior in albino rats. III. The effect of different levels of vitamin B₁ intake on convulsive seizures; the effect of other vitamins in the B-complex and mineral supplements on convulsive seizures. *J. comp. Psychol.*, 1942, **34**, 85-89.
134. PATTON, R. A., KARN, H. W., & LONGENECKER, H. E. Studies on the nutritional basis of abnormal behavior in albino rats. IV. Convulsive seizures associated with pyridoxine deficiency. *J. biol. Chem.*, 1944, **152**, 181-191.
135. PATTON, R. A., & LAZOVIC, A. D. Sensory pre-conditioning and the convulsions associated with magnesium deficiency in the rat. *J. comp. Psychol.*, 1946, **39**, 265-273.
136. PATTON, R. A., & LONGENECKER, H. E. Studies on the nutritional basis of abnormal behavior in albino rats. V. The effect of pyridoxine deficiency upon sound-induced magnesium tetany. *J. comp. Psychol.*, 1945, **38**, 319-334.
137. PAVLOV, I. P. *Conditioned reflexes*. (Trans. by G. V. Anrep.) London: Oxford Univ. Press, 1927.
138. PIKE, F. H., ELSBERG, C. A., MCCULLOCH, W. S., & CHAPPELL, M. N. The problem of localization in experimentally induced convulsions. *Arch. Neurol. Psychiat., Chicago*, 1930, **23**, 847-868.
139. POLLOCK, L. J. Experimental convulsions. *Arch. Neurol. Psychiat., Chicago*, 1923, **9**, 604-612.
140. PUTNAM, T. J., & MERRITT, H. H. Experimental determination of the anticonvulsant properties of some phenyl derivatives. *Science*, 1937, **85**, 525-526.
141. RIESS, B. F., & BERMAN, L. The mechanism of the insulin effect on abnormal behavior. *Amer. J. Psychiat.*, 1944, **100**, 674-680.
142. ROSEN, V. H., & GANTT, W. H. Effect of metrazol convulsions on conditioned reflexes in dogs. *Arch. Neurol. Psychiat., Chicago*, 1943, **50**, 8-17.
143. SACKS, J., & GLASER, N. M. Changes in susceptibility to the convulsant action of metrazol. *J. Pharmacol.*, 1941, **73**, 289-295.
144. SACKS, J., & MAIER, N. R. F. Studies of abnormal behavior in the rat. XI. Factors that influence the type of reaction to metrazol. *J. comp. Psychol.*, 1942, **34**, 331-340.
145. SACKS, J., MAIER, N. R. F., & GLASER, N. M. The influence of metrazol on the "neurotic pattern" in rats. *J. Pharmacol.*, 1941, **72**, 33-34.
146. SHARP, H. C., WINDER, C. L., & STONE, C. P. Effects of electroconvulsive shocks on "reasoning" ability in albino rats. *J. Psychol.*, 1946, **22**, 193-197.
147. SHEPLEY, W. H., & MCGREGOR, J. S. Electrically induced convulsions in treatment of mental disorders. *Brit. med. J.*, 1939, **2**, 1269-1271.
148. SHOHL, JANE. Effects of oral administration of dilantin sodium on abnormal behavior in the rat. *J. comp. Psychol.*, 1944, **37**, 243-250.

149. SIEGEL, P. S. The effect of electroshock convulsions on the acquisition of a simple running response in the rat. *J. comp. Psychol.*, 1943, **36**, 61-65.
150. SIEGEL, P. S., & LACEY, O. L. A further observation of electrically-induced "audiogenic" seizures in the rat. *J. comp. Psychol.*, 1946, **39**, 319-320.
151. SISK, H. L. The effect of experimentally induced audio-genic seizures upon relearning in the white rat. *J. Psychol.*, 1942, **14**, 85-88.
152. SISK, H. L. Maze learning ability and its relation to experimental audiogenic seizures in the rat. *J. gen. Psychol.*, 1944, **30**, 89-91.
153. SMITH, K. U. Quantitative analysis of the pattern of activity in audio-epileptic seizures in rats. *J. comp. Psychol.*, 1941, **32**, 311-328.
154. SNEE, T. J., TERRENCE, C. F., & CROWLEY, M. E. Drug facilitation of the audiogenic seizure. *J. Psychol.*, 1942, **13**, 223-227.
155. STAINBROOK, E. J. A note on induced convulsions in the rat. *J. Psychol.*, 1942, **13**, 337-342.
156. STAINBROOK, E. J. Maze behavior of the rat after electroshock convulsions. *J. exp. Psychol.*, 1943, **33**, 247-252.
157. STAINBROOK, E. J. Experimentally induced convulsive reactions of laboratory rats. I. A comparative study of the immediate reactions. *J. comp. Psychol.*, 1946, **39**, 243-264.
158. STAINBROOK, E. J. Shock therapy: psychologic theory and research. *Psychol. Bull.*, 1946, **43**, 21-60.
159. STAINBROOK, E. J., & DEJONG, H. Symptoms of experimental catatonia in the audiogenic and electroshock reactions of rats. *J. comp. Psychol.*, 1943, **36**, 75-78.
160. STAINBROOK, E. J., & LÖWENBACH, H. The reorientation and maze behavior of the rat after noise-fright and electroshock convulsions. *J. comp. Psychol.*, 1942, **34**, 293-299.
161. STONE, C. P. Physiological psychology. *Annu. Rev. Physiol.*, 1945, **7**, 623-652.
162. STONE, C. P. Effects of electroconvulsive shocks on daily activity of albino rats in revolving drums. *Proc. Soc. exp. Biol.*, N. Y., 1946, **61**, 150-151.
163. TSENG, F. Y. S. Differentiation of anger and fear in the emotional behavior of the rat. Unpublished Master's thesis, Univ. Toronto, 1942.
164. TSENG, F. Y. S. The differentiation of anger and fear in the emotional behavior of the rat. *Bull. Canad. psychol. Ass.*, 1942, **2**, 35-36.
165. TUFTS, ELMA V., & GREENBERG, D. M. The biochemistry of magnesium deficiency. I. Chemical changes resulting from magnesium deprivation. *J. biol. Chem.*, 1938, **122**, 693-714.
166. TURCHIOE, RITA. The effect of coramine on the facilitation of the audiogenic seizure. *J. comp. Psychol.*, 1945, **38**, 103-107.
167. TURNER, R. H. An approach to the problem of neurosis through the study of respiration, activity, and startle in the white rat as influenced by the difficulty of visual size discrimination. *J. comp. Psychol.*, 1941, **32**, 389-405.
168. WATSON, M. L. The inheritance of epilepsy and of waltzing in *Pero-myscus*. *Contr. Lab. Vertebr. Genet.*, Univ. Mich., 1939, No. 11. Pp. 24.
169. WEINER, H. M., & MORGAN, C. T. Effects of cortical lesions upon audiogenic seizures. *J. comp. Psychol.*, 1945, **38**, 199-208.
170. WINDER, C. L. The effect of electro-

- convulsive shock on general activity of rats. *Amer. Psychologist*, 1946, 1, 449.
171. WITTMAACK, K. Ueber Schädigung des Gehörs durch Schalleinwirkung. *Z. Ohrenhk.*, 1907, 54, 37-80.
172. WORTIS, S. B. Experimental convulsions. *Amer. J. Psychiat.*, 1932, 88, 611-621.
173. WORTIS, S. B. Experimental convulsive seizures. *J. nerv. ment. Dis.*, 1933, 77, 233-245.
174. WORTIS, S. B., SHASKAN, D., IMPASTATO, D., & ALMANI, R. Brain metabolism: VIII. The effects of electric shock and some newer drugs. *Amer. J. Psychiat.*, 1941, 98, 354-359.

A CRITICAL REVIEW OF THE LITERATURE ON "ABSOLUTE PITCH"

D. M. NEU

Indiana University

Pitch discrimination and, more specifically, "absolute pitch" have been a source of controversy and speculation for a great many years, beginning with Stumpf's discussion of Mozart's sense of pitch (43). In general, absolute pitch is the ability of an individual to name correctly a particular tone that is sung or played on an instrument, without comparing it to any other heard tone. The explanations of absolute pitch have been varied and many. The inadequacy of these explanations creates our present problem. Even more so than other forms of abilities, musical talent, especially absolute pitch, has traditionally been explained as an inborn quality or faculty that relatively few people are privileged to have.

In view of the fact that an inborn quality has been accepted so generally as an explanation of accurate pitch discrimination, one might hesitate to offer any other explanation if it were not for the fact that no real evidence has been presented to prove that there is such a quality. Moreover, if all the work on pitch discrimination is reviewed, while keeping in mind the real event taking place in pitch discrimination, the evidence points strongly away from any faculty or quality inherent in the individual. Rather, the evidence might show that the ability to discriminate pitch depends upon the life conditions of the individual and is probably built up during his lifetime.

It is the purpose of this paper to criticize the definitions, theories, and experiments relating to absolute pitch, from an interbehavioral standpoint (16). This point of view stresses the behavioral event taking place when an organism is interacting with a stimulus object in any particular situation. The stimulus value that the object has for the individual depends upon the individual's life history, the situation at the time, as well as upon previous interaction with the object. The response that the individual makes depends upon the stimulus value or function that the object has for him in any particular situation.

DEFINITION OF ABSOLUTE PITCH

The definition of the term "absolute pitch" has caused the greatest confusion in evaluating data from experimental work, because of the many different meanings of the term. As a result each investigator

criticizes every other investigator's work which yields results different from his own, by saying that the other investigator's subjects did not have absolute pitch or that the conditions allowed the subjects to use relative pitch, or any number of other similar reasons.

Petran (30) describes absolute pitch as follows:

... on the psychological side is an inherent assumption of the extensity and telephone theories. With this assumption in mind, judgments of absolute pitch may be defined as judgments based on association learned between more or less narrowly limited ranges of the pitch series, ... these judgments being without reference to or aid from any tone or tones recently heard which have been given as a standard or attended to in any degree as being of a certain pitch or familiar pitch position (30, p. 12).

Lewis (24) defines pitch as an attribute of auditory experience and not as a characteristic of acoustic waves.

Pitch is that attribute of auditory experience which determines the positions of sounds in a psychological continuum extending from low to high ... Pitch is no more a property of sound waves than color is a property of light waves. Pitch arises as an attribute of auditory experience when the ear is stimulated in appropriate ways, just as color arises as an attribute of visual experience when the eye is appropriately stimulated (24, p. 121).

The performance in absolute pitch is defined by most writers as designating a heard tone correctly from memory alone, without any other aid. In the opinion of the present writer, absolute pitch might be better defined as the ability to discriminate tones without the aid of other tones to such a degree that naming or pointing out the note is rarely incorrect.

Bachem (3) probably has made the most complete attempt to categorize ways of discriminating pitch. Genuine absolute pitch is based upon immediate recognition of the chroma, which would be the same definition most investigators would give to it. Quasi-absolute pitch, he says, is based upon aural or vocal standard and the interval sense. Pseudo-absolute pitch is based upon the estimation of tone height. Bachem's types of pitch would probably be agreed upon by most writers. He has only given names to the various degrees of pitch discrimination. However, it is only too obvious from the literature that many investigators have taken what Bachem calls quasi- or pseudo-pitch in judging genuine absolute pitch. For this reason a very reliable collection of data from absolute pitch experiments cannot be made.

Bachem says that the outstanding difference between the "absolute" pitch of most investigators and his type of genuine absolute pitch is the accuracy of judgment. The average error of judgment in cases of

"pseudo-absolute" pitch amounts to about five half-tones (after training) and in "genuine-absolute" pitch it is one-eleventh of a half-tone according to Abraham, and even less according to other authors. We would agree with Bachem and others that absolute pitch discrimination is a matter of accuracy of judgment. This evidence furthers all the more the explanation that absolute pitch is merely discriminatory ability developed to a very fine degree. Ability here does not imply some special gift which the individual has inherited, but a development of auditory discrimination to the greatest accuracy during the lifetime of the individual, based upon his reactional biography. Thus, absolute pitch is only the most accurate degree of pitch discrimination.

Although Seashore is probably the foremost proponent today of absolute pitch as a mysterious gift or power, he unknowingly substantiates our view by stating:

As to the degree of accuracy of free tonal memory, we find a gradual transition from the very finest absolute pitch through all degrees of accuracy down to crude forms of acquired pitch. But it is convenient to make the distinction that absolute pitch may be thought of as involving accuracy to a very small fraction of a semitone; whereas acquired pitch is usually thought of in no finer terms than semitones (37, p. 18).

THEORIES OF ABSOLUTE PITCH

Practically all explanations of absolute pitch assume an inborn quality of some sort, and in some degree, although the trend in recent years has been to regard the ability as an acquisition. Probably the best known theories of absolute pitch are the so-called physiological theories. However, we shall not consider physiological theories in this paper for we strongly maintain that the stimulus object and situational factors involved must be considered, as well as the individual, in any explanation of behavior. The above does not mean that we are ignoring the important work that has been done by such men as Stevens (40, 41, 42), Thurlow (44), Wever (48), Wever and Bray (49, 50, 51), Wilkinson and Gray (53), and others upon differential "place" responses in the cochlea, but we maintain that even if such differential physiological reactions take place, they do not adequately explain pitch discrimination because they do not represent the entire behavioral event which occurs.

Physiological and physical data must not be confused with psychological behavior. Since we are interested in the psychological or behavioral event, that is, pitch discrimination, it is necessary for us to state what this psychological event consists of and thereby point out the futility of most of these theories, and their inability ever to give a

satisfactory explanation. In the act of discriminating a "pitch" we are interested in the particular individual determining a tone. That is, the tone is the stimulus and the factors that operate in the interaction between the individual and the tone are our problem. It is clear then that the theories based only upon physiological or physical data are not considering the actual psychological event, but rather only a part of it, the medium of contact. When we use the term, media of contact, we are speaking of the means by which the individual is able to react to the object. For example, in the case of pitch the individual is able to react to a tone because a medium of waves is present and they enable the individual to make contact with the sound of the stimulus object and to respond to it. The sound waves, then, are not the stimulus object but only the medium of contact between the individual and the stimulus object.

Practically all views of absolute pitch involve one of two explanations:

1. The individual may possess a special ability, differing from anything possessed by the average individual.
2. The individual may possess a high degree of some ability which is found to some extent in all persons.

Boggs (8), Révész (33), and Baird (5) are exponents of the first view. They believe that *quality* is the distinguishing feature of a tone in absolute pitch judgments. Boggs was at first inclined to believe that ability to hear overtones was the cause of keen pitch discrimination, but later changed her view from timbre to quality. Quality, she says, enables a person to hear tones in much the same way as those with color vision see colors. Baird's view is similar. He bases it chiefly on octave-errors and on immediateness of judgment, like the recognition of red when once it has been pointed out as red. Révész makes a definite distinction between those who recognize tones by pitch, and those who recognize them by quality. Both are supposed to have genuine absolute ear but the latter group has judgments more immediate and more correct. Révész holds that ability to distinguish quality is innate, like colors in vision, and cannot be acquired. Pitch-memory, on the other hand, may be acquired and perfected by diligent practice. To explain the influence of timbre, he says it is possible that when other parts of a tonal complex are changed the quality cannot be recognized in the new combination.

For von Kries (20), absolute pitch is something which cannot be acquired, while Stumpf (43) leaves the question open and gives more weight to familiarity, musical ability, feeling, and interest as displayed

in practical musical activities. Mull says, "The most convincing proof of the spuriousness of tonality, in Baird's sense of the term, would be success in training average individuals in absolute pitch to the point where they are similar to the 'gifted' individuals in every respect" (28, p. 475).

Seashore says, "Absolute pitch is an inborn predisposition which manifests itself during childhood and is strikingly immediate and spontaneous, . . ." (35, p. 18). He believes that absolute pitch involves errors of a small fraction of a semitone, whereas acquired pitch operates in terms of steps in the musical scale. Seashore considers that both are forms of memory and that both rest upon a favorable hereditary capacity, although both are influenced by training. But absolute pitch is influenced only to a slight extent, while acquired pitch can be learned to some degree by any person with a good sense of pitch and it is dependent upon knowledge of musical steps and notes.

Explanations of absolute pitch which do not rest upon a special ability are held by Meyer (27), Köhler (18), Watt (45), and Ogden (29).

Meyer (27) says that if we are to assume that some people are gifted with absolute pitch, it means that there is a physiological property, the lack of which prevents the acquiring of this mental faculty. It would be difficult to say what the physiological property would be. Rather than to make such an assumption, Meyer favors the idea that everyone potentially has the ability. He says that everyone has a certain amount of memory of pitch, since they are able to recognize and discriminate. Meyer also states that there is no reason proved by any experiment which shows that the individual by systematic and sufficiently lasting practice may not be trained to better "absolute pitch." He considers some features of the tone-body, rather than pitch, to be the basis of absolute pitch judgments.

The theory that tone-body, and not pitch, is important has been developed by Köhler (18). Tone-body includes brightness, vocality, volume, and intensity. It is recognized immediately without the intervention of a memory image. Watt (45) disagrees with Köhler's list of attributes and considers that Köhler has merely shown that timbre is important. Watt believes the attributes of tone to be two, pitch and volume. Both are taken into account by absolute ear, where there may be either a special refinement of sensory apparatus or a better auditory memory. But he thinks that in many cases of absolute pitch there is required also an absolute point of reference in auditory pitches. The problem is analogous to that of localization of the skin, where, however, the great influence of relativity does not exist. It is due to the lack of

relativity that tactual localization is possible for all persons. Auditory localization is possible only to a few who have not allowed the relative ear to dominate the absolute ear. According to this view, we all have had a chance to possess absolute pitch, but because of our education in musical relationships we have lost it through disuse.

Ogden (29) bases absolute ear upon a refined sensitivity to different clang-patterns and also insists upon the necessity of reference to a scale with which the individual who possesses the talent is familiar. This explanation is essentially like that of Watts. With Ogden, however, the point of reference seems to be introduced for purely systematic reasons and to be considerably overemphasized, especially when one remembers that good interval-judgment is not a requisite for absolute pitch.

Some investigators think that movement of the muscles in the throat are a help in making a judgment of pitch. Lotze (26) and Köhler (18) thought there must be a "larynx sensation" and thus pitch discrimination was limited to those with singing voices. Ladd and Woodworth (22) thought an innervation of the larynx is present when one is judging a note. Gough (13) in an extensive experiment could find no evidence of this. Wyatt makes this statement:

... it seems probable that proficiency in a pitch discrimination test may be affected by many factors which are quite remote from simple auditory sensation and which may involve, not just the auditory sensorium, but possibly even the entire organism (58, p. 54).

Lewis has progressed farther than most investigators in attributing the pitch discrimination of an individual to that particular individual's experience. However, he is holding to traditional ideas for he says: "The sense of pitch, an innate capacity . . ." (23, p. 349). But then again, he also says: "Pitch is not a hard and fast thing like a physical quantity. To regard pitch as something rigid and fixed is to overlook the basic meaning of tonal experience" (23, p. 349). Lewis' main point is that pitch is not to be confused with frequency, which most of the older writers have made one and same (30).

Bachem (4) thinks inheritance, attention, and experience are the factors in the creation of absolute pitch. His factors present a more objective approach, especially if by inheritance is meant only the biological make-up of a human being. Attention is important for the individual must attend to or contact the stimulus object in order to interact with it, and experience is the reactional biography or life history that is more important in determining the individual's behavior (16).

Even though the interactional history with the tone stimulus may be subtle, it is clear that the reactions of the individual would have better opportunity to be highly developed if interaction with the tone were deliberately undertaken. We mean by this statement that the discrimination of pitch, even when it is developed to a point where it is "absolute," is developed during the lifetime of the individual. Absolute pitch is a matter of the keen development of attending to the stimulus (the tone) and the keen development of perceiving the stimulus (differentiating the tones).

If we are to consider that pitch discrimination is developed not from some inborn power but during the lifetime of the individual's reactions, it is not important whether the investigator elicits *pseudo*, *quasi*, or *genuine* absolute pitch in his experiment. It is not important because, from this interbehavioral view, the individual's accuracy of discrimination reflects only the degree to which he has interacted with tones during his lifetime. The individual with discrimination of the kind that we have chosen to call "genuine" absolute pitch has obtained a psychological interaction with different tones equivalent to what most of us have with the various colors. This does not mean necessarily that the individual has to "train" himself or to practice deliberately with tones in order to judge them. The interaction with the tone stimulus could be as subtle as any psychological interaction can be; for example, as subtle as how we learn to distinguish red from green.

That there are usually great differences between those who "have" absolute pitch and those who are "trained" along that line is not the crucial point which most writers seem to think. The point that needs to be emphasized is the uselessness of merely attributing absolute pitch to some unknown power. If we are to find a scientific explanation, we must deal with the actual event of the tone and the organism discriminating it. With an interbehavioral explanation we can deal with the facts as they are present in the crude event (16), viz: the discriminating of pitch so accurately that it is "absolute" and the learning of this ability during the lifetime of the individual. This means that we will have to consider the individual's reaction from the beginning of his psychological interaction with tones of any sort. How the person will develop his reactions to tones depends upon the individual in his life conditions. Therefore, the individual who cannot discriminate pitch after training as well as the individual with absolute pitch but without special training is not able to overcome the effects of reaction to tones that developed during his lifetime before training was started. We know that the first years mark the beginning of basic permanent behav-

ior, and so the older the person the more difficult it becomes for him to be influenced by training (16).

INVESTIGATIONS OF ABSOLUTE PITCH

Some main points in the results of experiments that Petran (30) reviewed up to 1932 were that white notes on the piano are more often correctly judged than black; that notes in the middle range are judged more accurately than those in the extreme range; and that there is a tendency to judge high notes lower and low notes higher than they actually are. In a recent experiment Riker (35) found that subjects with musical experience were able to judge pitch better in the middle range. All of these results show better judgment with notes that in general are used most often. In other words, notes that are played more often and experienced more often by the individual are the ones he knows best.

Most investigators did attempt to consider experience as an explanation, for one of the most frequent explanations offered was that the more familiar a timbre, the easier it was to judge. However, von Kries (20) and Abraham (2) both rejected this explanation. Stumpf (43) and Köhler (18) both concluded that experience must be considered. Chilloff (10) concluded that absolute pitch depends on the faculty of determining the timbre of sounds of different heights. Baird (5) found subjects trained in music to be less accurate with tones on other instruments than on a piano. Köhler (18) holds that unfamiliar timbre recognition is quite impossible. Gough (13) after training both musical and unmusical subjects found only comparatively slight differences in accuracy of judgment with different timbres. Thus, from these results it seems that familiarity of timbre is important in judgment of pitch. Further results pointing to familiarity are the experiments by Stumpf (43) who found that a bass player was better in judging tones in the lower range than in the rest of the scale and that a violinist was better in judgment in the upper range.

Association with tones. Association with tones was reported by many observers such as Stumpf (43), Whipple (52), Boggs (8), and Gough (13). One was a keyboard association so that the subject had a spatial position of the note. Another was an association with notes of a prominent song or triad built on the tone. A non-musical association was with pain, such as with high notes, or tones sounding woody or muffled to the subject. All these results suggest that past experience is necessary before the associations with which to make judgments can occur.

Practice effects. Practically all of those individuals with absolute

pitch are musicians of more than average competency (28). Bartholomew (6) found that proficiency of subjects increased with the number of ear training and harmony courses they had. Jadassohn before 1899 stated that "the faculty of absolute pitch was easily acquired" (15, p. 33). The first experiment on acquiring absolute pitch was made by Meyer (27) on himself and another person. They learned to name tuning forks from 100 to 4500 vibrations per minute by their vibration numbers, and thirty-nine notes on the piano, with a rather high degree of accuracy. They began with a few pitches at a time and added new ones from time to time. They did this with tuning forks and a piano. With a choice of 39 pitches, more than half of their judgments were correct. After several years the subjects lost the greater part of what they learned for want of continued practice. Boggs (8), experimenting on herself, improved her judgment by paying special attention to overtones; and Köhler (17) did the same thing by concentrating on what he calls "body-tones."

Both Gough (13) and Mull (27), experimenting with college students, found that a certain amount of pitch naming can be acquired by musical and unmusical subjects. The two experiments show that systematic practice in pitch identification is effective.

The results of Mull's experiment (28) show that ability to judge notes correctly can be greatly improved by training. The improvement is largely immediate and is relatively lasting in its effects. Mull states that in making judgment of absolute pitch a high degree of attention to the notes is more effective than a lower degree, and a note which is not attended to has no noticeable effect. She also finds that the ability to acquire absolute pitch correlates better with practical musical ability than with ability in the Seashore tests. Mull concludes that the average person can acquire absolute pitch. She bases this statement on the fact that there is a similarity of performance between Baird's "gifted" subjects and Mull's group of average persons after extensive training. Mull goes on to state that the possession of absolute ear seems to rest upon an interest in the notes themselves rather than in their melodic or harmonic relationships. The importance of attention is indicated by (a) the immediateness and permanence of the learning, (b) the individual variability from day to day, (c) the fact that a stimulus which is not attended to has no noticeable effect, and (d) the fact that a good hearing of the note is necessary before judgment can be made.

But, as Mull says, "There still remains the question why certain individuals without having been trained possess absolute ear so much more completely than others." Mull's answer to this question presents

one possibility for the development of absolute pitch and is certainly more feasible than attributing it to an inborn quality. She says: "An answer may lie in the suggestion that such persons, probably when they were children, found notes interesting in themselves . . ." (28, p. 492).

The percentage of correctness among "gifted" persons in Baird's study (7) ranged from 100 per cent to 0 per cent, both by group and by individuals. Boggs (8) percentage of correct judgments in his "gifted" subjects ranged from 100 per cent to 16 per cent in the same individual. Mull (28) thinks it depends upon experimental conditions and upon the individual himself.

Most investigators found the middle pitches to yield the greatest precision in judgment, but this was not always the case. Stumpf (43) recognized that the greatest precision is in the middle region and that pitch discrimination is finest there. But he attributes it to greater use and consequent familiarity, for we have already mentioned his subjects, the bass player and the violinist, who discriminated best at the lower and higher ranges, respectively.

Riemann (34) said that some musicians have absolute pitch because of their frequent tuning of their instruments. Wilson (55) agrees with this. Wyatt (58), in one of the most recent experiments, showed that pitch discrimination of initially "pitch deficient" adults was significantly improved after intensive training. She places special emphasis on the need for the right kind of training in order to get improvement in pitch discrimination.

Another argument for the sufficiency of an explanation on the basis of previous experience and familiarity is the fact that many investigators have found some order of difficulty among various instruments. Von Kries (20) finds general agreement among musicians that the order from least to most difficult is: piano, strings, wind instruments, voice, whistling, tuning forks, and bells. Baird's order (18) is piano, organ, voice, and tuning forks. Mull (28) points out that it is possible that this could be due to familiarity, but Baird (7) and von Kries (20) do not care to believe it. The fact that Baird and von Kries have a slightly different order shows that it depends upon the particular individual's experience as to what order he would choose. A piano player might be more likely to judge tones correctly on any musical instrument rather than on a tuning fork or bells with which he would have little contact. More specifically, a violinist may be able to judge tones more accurately than a pianist because it is necessary for him to play tones on his instrument by placing his finger in the correct position; but a pianist has

merely to touch a key which hits a string already tuned to a particular tone. Bachem (4) substantiates this by his statements:

Musical experience, however, plays a role in the acuity of absolute pitch. . . . Several (subjects) believed that musical experience improved their quality of absolute pitch, and lack of experience allowed it to decline. . . . Two musicians found their absolute pitch declining with increasing deafness, which had not progressed sufficiently, however, to interfere with their musical activities. . . . A somewhat greater acuity of absolute pitch was found in violinists than in piano players (4, p. 439).

Cameron (9) conducted an experiment to see the effects of practice in discrimination and singing of tones. He found no marked correlation between the initial capacities of the subjects tested for discrimination of tones and ability to reproduce these tones accurately by singing. Practice in singing tones of a certain pitch resulted in marked reduction in the error of reproducing those tones in the case of four of six subjects. Slight improvement in singing tones of a pitch different from the one practiced was made by three subjects and no improvement in the other. Subjects who improved in accuracy of singing tones of a certain pitch improved also in discrimination of tones at that level. There was no improvement for subjects in the discrimination of tones of a different pitch from that practiced. Subjects who did not improve in accuracy of singing made no improvement in discrimination.

Lewis (25) conducted an experiment to secure data for constructing and comparing pitch scales. The scales he constructed did not agree with each other nor with scales based on the findings of Shower and Biddulph (39) and of Stevens and Volkmann (42). The difference could not be attributed to experimental error. Lewis concludes that any pitch scale depends, in some measure, upon the group of persons who made the required judgments and upon the particular method of observation employed.

If it is true that pitch depends upon the particular individual's judgment, then each judgment would not necessarily be alike, as Lewis' investigation shows. This result certainly is not evidence that pitch is an inborn quality, for if so, then a certain number of vibrations of a tone would be heard as exactly the same pitch judgment by every individual. Rather, it is evidence that the ability to discriminate pitch is acquired in the person's lifetime, and that accuracy in hearing a pitch as a certain note depends upon his own reactional biography.

Bachem (4), in one of the most recent experiments, studied 103 cases from which he devised his three types of absolute pitch discrimination:

genuine, *quasi*, and *pseudo*. All of these types of absolute pitch discrimination occur with variation in acuity, range of the musical scale, and influence of timbre. These variations occur because of differences in past experience and in familiarity with tones.

Bachem in an investigation to secure comparative results conducted identical experiments on persons with and without absolute pitch. A first pair of experiments was performed on the piano by offering a tone to a person placed back toward the piano and then having the person turn toward the piano and identify the key without striking it. Twenty tones were judged this way and another twenty were identified by striking the key. In order to compare the results of this with accuracy of absolute pitch judgment, several ranges of the piano scale were tuned to eight tones, and marked correspondingly. Most subjects identified the tones correctly.

A second pair of experiments was designed for more accurate results. A tone was offered several times on a tone variator which was standardized in vibrations per second. Every five minutes a different tone was sounded and altered according to directions of the subject, until the tone was considered the right one. A similar experiment was conducted with the five-minute intervals replaced by half-day intervals.

The average error of absolute pitch determination was 0.6 per cent for observations extending over one-half hour. For persons without absolute pitch, the error amounted to about 8 per cent on the tone variator, representing an interval of one to two half-tones. On the piano an error of about five half-tones was found. With these results Bachem says there is conclusive evidence of the great difference in accuracy of pitch identification in persons without and with absolute pitch.

As Bachem explains, attention, or what the interbehaviorist would call attending to stimuli (16), is an important factor to be considered. Attending to stimuli we must consider as having the same sort of chance for development as reacting to tones, given in the explanation above. An obvious example is the development of hearing of a blind person. Bachem (4) tells of a blind person without musical training who detected his absolute pitch accidentally during a conversation with a possessor of absolute pitch and he was able immediately to identify tones correctly after he knew how to denote them. Bachem admits that the large percentage and high degree of absolute pitch in congenitally blind persons shows that inheritance is not the only factor. Eleven of his 103 cases were blind since birth and in none of these cases were there any relatives who had absolute pitch. Bachem also admits that there is an influence of forced attention to sound by the blind.

Regardless of such evidence, Bachem shows that he is influenced by

traditional views for he thinks that heredity plays some part. "An alternate possibility," he says, "could be that the musical environment surrounding these families played an important role" (4, p. 437). He rules this out in the case of two families that he studied, because in one the grandfather had absolute pitch and the mother was not very musical. In another, two musical prodigies studied on a piano which was so much out of tune that they had to distinguish between the name of the key and the actual tone produced. Considering the first case, there is no indication that the child had not associated with the grandfather and also according to an interbehavioral explanation, it would not make any difference if the child had not associated with anyone who had absolute pitch. In regard to the second case, the conditions under which the children worked would mean that they would have to attend to tones even more carefully than most students, and this would help, not hinder, the acquisition of absolute pitch.

It is true, of course, that a musical environment may help the development of pitch discrimination because of the added opportunity to attend to tones. Absolute pitch occurs at a decidedly higher percentage in musical groups, particularly professional musicians. Bachem states, "Especially is this true of those with early musical training." He goes on to say, "It seems that attention to musical tones in early youth plays a predominant role in the development of absolute pitch" (4, p. 439).

Besides the statements made by these investigators, the results of many other experiments show the incorrectness of saying that absolute pitch is an absolute inborn quality. Meyer and Heyfelder (27) attempted to train themselves and their results show a high degree of correct judgments. Köhler (18) after two weeks of practice could make 112 correct judgments out of 220. Gough (13) trained 90 students and got an improvement. Mull's data (28) show that practice increases the percentage of correct judgments from 40 per cent to 82 per cent. Wedell states, "It is evident that some of the later experimenters are aware that even in absolute pitch, the judgment is at least in part one of relation" (46, p. 488).

Wedell's experiment led to these conclusions:

1. Relatively unmusical observers can increase their accuracy.
2. Greatest increase is in the first few practice sessions.
3. Limit of ability is an average error of about three semitones.
4. The course of the learning process is very irregular and there are large individual differences.
5. Unmusical subjects can learn accurately and easily to recognize tones that are eight and one-half semitones, but they fail to judge intervals when decreased to five and one-half semitones or less.

Riker (35), in an experiment comparing subjects who said they had pitch-judging ability with those who disclaimed it, concluded that ability to judge pitch was not confined to specially trained or talented subjects. He also concluded that there were various degrees of accuracy in judging pitch and that it was "... a function of day to day experience with music" (35, p. 346).

Absolute pitch in children. Very little study has been made on children by early investigators, but Abraham (1) trained three four-year-old girls to sing "Ade" to the mode a-d. One had been to kindergarten where songs were transposed and she sang it at various pitches but with the correct interval; the other two were always correct even after three months and one after one and three-quarter years.

Bennedik (7) had three children who all acquired absolute pitch ability but in different measures. Copp (11) taught children to name heard tones. From her experience with school children she states that 80 per cent of all children can acquire absolute pitch. Katz (17) and Bennedik (7) substantiate this but not with as high a percentage. Komatsu (19) experimented with a child, four years and ten months at the beginning of training, who was able, after one year, to discriminate 46 harmonics consisting of three or four tones.

A number of experiments as to whether absolute pitch can be acquired by anyone with normal hearing who wants to learn have been performed using tuning forks, piano, and oscillators (34). All experiments point to the conclusion that pitch naming ability can be improved to a certain extent by training. Since the learning period seldom covered more than a few months no greater result could be expected.

Probably the most significant and thorough experiment was conducted by Wolner and Pyle (56) to show what effect age and practice have upon pitch discrimination. Music teachers in three elementary schools selected pupils showing the greatest deficiency of pitch discrimination. From this group the seven poorest were selected. No pupil selected could distinguish the thirty-vibration differences on the forks. On the piano, in general, they could not distinguish differences of the octave, fifth, third, whole tone, or semitone. None of them could sing, although they had been in music classes from the first grade and were at the time of the experiment in the fifth, sixth, and seventh grades. There were three boys and four girls.

Each pupil received individual instruction and tests for twenty minutes each morning, five days a week. The whole number of hours spent in training was on the average 16 for each pupil, extended over a period of 81 days.

The definition and meaning of pitch, or "high" and "low," as distinguished from intensity, duration and timbre were strongly and repeatedly emphasized, particularly in the early days of training. The pupils were led to see the necessity of thinking of tones as one would think of a problem. Interest, attention, and concentration were worked for in the method of teaching.

All seven pupils learned to discriminate perfectly the intervals of octaves, fifths, wholetones, and semitones in the range from A, sixteen tones below middle C, through A, thirty-four tones above middle C, a tonal range of four octaves. There was great variability in the time required to reach a given degree of efficiency, and in the response to methods and changes of methods. With the forks, four of the pupils became perfect in distinguishing all the pitch differences from the largest, thirty vibrations, down to the smallest difference, one-half a vibration. Of the other three pupils, one learned to distinguish perfectly down to two vibrations difference, another down to three vibrations difference, and the third down to five vibrations difference. The standard of perfection was 10 out of 10 trials correct.

Each pupil improved noticeably in ability to sing. At the conclusion of the experiment, one pupil sang the words and music of several songs, with no trace of pitch deficiency; and also sang major and minor scales, chromatics, intervals, and tones picked at random. Another sang scales and intervals and the music of a song without words. Two pupils sang scales and intervals. The other three pupils sang, not perfectly, but with tremendous improvement over their initial efforts.

CONCLUSIONS

The statements made by the writers we have quoted and the results of the many experiments we have discussed certainly show the need for an adequate explanation of absolute pitch. The results are:

1. Absolute pitch and lesser degrees of pitch discrimination can be acquired by some individuals. Thus, it proves false the idea that absolute pitch is some rare power over which we have no control.
2. Behavior that closely entails pitch discrimination, such as musical experience and musical training, allows a much better opportunity to develop keener pitch discrimination.
3. Pitch discrimination can be acquired readily in early age, which may be an indication that the younger the reactional biography the easier it is to acquire new reactions.
4. The development of keener attending to stimuli makes for a keener development of pitch discrimination. Obvious examples of this are congenital blind persons and child prodigies.
5. The subtlety of behavioral development is brought out by those indivi-

duals who have built up in their lifetime keen pitch discrimination and suddenly realize that they are able to perform such behavior.

We can see that an interbehavioral view is justified by the results of past experiments. Basically, the interbehavioral explanation of absolute pitch rejects the conception of an inherent faculty or quality. It accounts for absolute pitch and all pitch discrimination as behavior developed within the lifetime of the individual. This means that the discrimination is due to the way in which the individual builds up reactions to sounds, and more specifically to the thoroughness in his development of attending to tones (stimuli). Accordingly, the individual's reactional biography, built up from past experience, is the important factor in determining what sort of tonal discriminations he will make. It follows then that "absolute pitch" is nothing more than a fine degree of accuracy of pitch discrimination. The pitch of a tone is the sound that the particular individual learns as that particular tone.

BIBLIOGRAPHY

1. ABRAHAM, O. Das absolute Tonbewusstsein. *Sammelbände d. Internat. Musikgesellschaft.*, 1901, 3, 1-86.
2. ABRAHAM, O. Das absolute Tonbewusstsein und die Musik. *Sammelbände d. Internat. Musikgesellschaft.*, 1907, 8, 486-491.
3. BACHEM, A. Various types of absolute pitch. *J. acoust. Soc. Amer.*, 1937-38, 9, 146-151.
4. BACHEM, A. Genesis of absolute pitch. *J. acoust. Soc. Amer.*, 1939-40, 11, 434-439.
5. BAIRD, J. W. Memory for absolute pitch. In *Studies in psychology, Titchener commemoration volume*. Worcester: Wilson, 1917.
6. BARTHOLOMEW, W. *A study of absolute pitch ability*. Master's thesis, George Washington Univ., 1925, 1-23.
7. BENNEDIK, F. *Die psychologischen Grundlagen der musikalischen Gehörbildung mit Beziehung auf die pädagogische Bedeutung der Tonwortmethode von Eitz*. Langensalza: Beltz, 1914.
8. BOGGS, LUCINDA P. Studies in absolute pitch. *Amer. J. Psychol.*, 1907, 18, 194-205.
9. CAMERON, E. H. Effects of practice in the discrimination and singing of tones. *Psychol. Rev.*, 1917, 23, 159-180.
10. CHILOFF, C. L. Des éléments de l'ouïe absolue. *Acta otolaryngol.*, 1930, 14, 382-392.
11. COPP, E. F. Musical ability. *J. Hered.*, 1916, 7, 297-305.
12. FLETCHER, H. A. Theory of hearing. *J. acoust. Soc. Amer.*, 1930, 1, 311-343.
13. GOUGH, EVELYN. The effects of practice on judgment of absolute pitch. *Arch. Psychol., N. Y.*, 1922, 7, No. 47.
14. HEIN, H. Über die Möglichkeit eines allgemeinen latenten absoluten Tonbewusstseins. *Z. Musikwissenschaft.*, 1929, 11, 414-419.
15. JADASSOHN, S. *A practical course in ear training*. (Translated by Campbell.) Leipzig: Breitkopf u. Härtel, 1899.
16. KANTOR, J. R. *Survey of the science of psychology*. Bloomington, Ind.: Principia Press, 1933.

17. KATZ, D. Über einige Versuche im Anschluss an die Tonwortmethode von Karl Eitz. Kongress für exper. Psychol. Bericht 6. Leipzig: Barth, 1914.
18. KÖHLER, W. Akustische Untersuchungen III. *Z. Psychol.*, 1915, 72, 159-177.
19. KOMATSU, A. Experiment on training of discrimination in absolute pitch. *Kyoiku Shinri Kenkyu*, 1940, 15, 203-205.
20. KRIES, J. v. Über das absolute Gehör. *Z. Psychol.*, 1892, 3, 257-279.
21. KRIES, J. v. *Wer ist musikalisch?* Berlin: Springer, 1926, 32-37.
22. LADD, G. T., & WOODWORTH, R. S. *Elements of physiological psychology*. New York: Scribners, 1911.
23. LEWIS, D. Pitch, its definition and physical determinants. *Univ. Ia. Stud. Psychol. Music*, 1937, 4, 346-373.
24. LEWIS, D. Pitch as a psychological phenomenon. *Music Teach. nat. Ass. Proc.*, 1939, 34, 121-133.
25. LEWIS, D. Pitch scales. *J. acoust. Soc. Amer.*, 1942, 14, 127.
26. LOTZE, H. *Medizinische Psychologie*. Leipzig: Weidmann, 1852.
27. MEYER, M. Is the memory of absolute pitch capable of development by training? *Psychol. Rev.*, 1899, 6, 514-516.
28. MULL, HELEN K. Acquisition of absolute pitch. *Amer. J. Psychol.*, 1925, 36, 469-493.
29. OGDEN, R. M. *Hearing*. New York: Harcourt, Brace, 1924.
30. PETRAN, L. A. Experimental study of pitch recognition. *Psychol. Monogr.*, 1932, 42, No. 6.
31. PETRAN, L. A. Nature and meaning of absolute pitch. *Music Teach. nat. Ass. Proc.*, 1939, 34, 144-152.
32. RÉVÉSZ, G. Über die beiden Arten des absoluten Gehörs. *Z. Internat. Musikgesellschaft.*, 1913, 14, 130-137.
33. RÉVÉSZ, G. *Zur Grundlegung der Tonpsychologie*. Leipzig: Veit, 1913.
34. RIEMANN, L. Das absolute Gehör. *Neue Musik-Zeitung*, 1908, 29, 515-516.
35. RIKER, B. L. The ability to judge pitch. *J. exp. Psychol.*, 1946, 36, 331-346.
36. SEASHORE, C. E. Inheritance of musical talent. *Mus. Quart.*, 1920, 6, 586-589.
37. SEASHORE, C. E. Acquired or absolute pitch. *Music Educators J.*, 1940, 26, 18.
38. SEASHORE, C. E. Quality of a musical tone. *Psychol. Bull.*, 1941, 38, 548.
39. SHOWER, E. G., & BIDDULPH, R. Differential pitch sensitivity of the ear. *J. acoust. Soc. Amer.*, 1931, 3, 275-287.
40. STEVENS, S. S., DAVIS, H., & LURIE, M. H. Localization of pitch perception on the basilar membrane. *J. gen. Psychol.*, 1935, 13, 297-315.
41. STEVENS, S. S., & DAVIS, H. *Hearing, its psychology and physiology*. New York: Wiley, 1938.
42. STEVENS, S. S., & VOLKMANN, J. Relation of pitch to frequency. *Amer. J. Psychol.*, 1940, 53, 329-352.
43. STUMPF, C. *Tonpsychologie*. Leipzig: Hirzel, 1883.
44. THURLOW, W. R. Binaural interaction and the perception of pitch. *J. exp. Psychol.*, 1943, 32, 17-36.
45. WATT, H. J. *The psychology of sound*. Cambridge, 1917.
46. WEDELL, C. H. Nature of absolute judgment of pitch. *J. exp. Psychol.*, 1934, 17, 485-503.
47. WEDELL, C. H. A study of absolute pitch. *Psychol. Bull.*, 1941, 38, 547-548.
48. WEVER, E. G. The electrical responses of the ear. *Psychol. Bull.*, 1939, 36, 143-186.
49. WEVER, E. G., & BRAY, C. W. Action currents in the auditory nerve in

- response to acoustical stimulation. *Proc. nat. Acad. Sci., Wash.*, 1930, 16, 344-350.
50. WEVER, E. G., & BRAY, C. W. The perception of low tones and the resonance-volley theory. *J. Psychol.* 1936, 3, 101-114.
 51. WEVER, E. G., & BRAY, C. W. Hearing in the pigeon, as studied by the electrical responses of the inner ear. *J. comp. Psychol.*, 1936, 22, 353-364.
 52. WHIPPLE, G. M. Studies in pitch discrimination. *Amer. J. Psychol.*, 1903, 14, 289-309.
 53. WILKINSON, G., & GRAY, A. A. *The mechanism of the cochlea*. London: Macmillan, 1924.
 54. WILLGOOSE, F. L. Absolute pitch and its attainment. *Etude*, 1929, 47, 144.
 55. WILSON, C. W. The gift of absolute pitch. *Musical Opinion & Musical Trade Rev.*, 1911, 34, 753-754.
 56. WOLNER, M., & PYLE, W. H. An experiment in individual training of pitch-deficient children. *J. educ. Psychol.*, 1933, 24, 602-608.
 57. WUNDERLICK, H. Four theories of tonality. *J. Musicol.*, 1941, 2, 171-180.
 58. WYATT, RUTH F. Improvability of pitch discrimination. *Psychol. Monogr.*, 1945, 58, No. 2.

SELECTIVE SAMPLING IN PSYCHOLOGICAL RESEARCH

ELI S. MARKS

National Office of Vital Statistics

Through a long process of indoctrination, psychologists have been made fully aware of the importance of statistical tests of significance in the analysis of data. They are, unfortunately, less aware of some of the basic requirements underlying these tests.

Some thought has been given by research workers in psychology to the importance of the form of distribution in the population sampled. Much less attention has been given to the even more important question of the method used in drawing a sample. This article presents two examples of difficulties in the interpretation of data arising from methods of sample selection fairly common in psychological research (and in research in other fields).

This article is not intended as a criticism of individuals. Residents of glass houses should avoid rock-heaving, and the writer's own research work has suffered notably from the very defects stressed in the present analysis. The writer cannot even promise not to repeat past errors since practical considerations frequently defeat the best of intentions. It is important, however, that all of us realize the limitations of our techniques and not claim for our conclusions a validity not warranted by our methods.

The theoretical consequences of selective sampling are well established and have been cited frequently. Nevertheless, this theory has been ignored fairly consistently in psychological research. All of us have been, to a greater or lesser extent, responsible for this situation. The examples selected for this article are not unique nor even unusual. The two sets of data analysed herein were selected partly from considerations of ease in analysis and partly because they are drawn from studies which were carefully planned and executed. The writer feels that the results of these two studies are generally valid but that it is necessary to point out that this validity cannot be conclusively established by the sampling techniques used.

A fundamental principle of sampling is that random selection must be present at some point in the sampling process. Without random selection it is not possible to determine whether the results obtained represent basic characteristics of the population or whether they are by-products of the sampling process. This fundamental consideration has not been sufficiently stressed in psychological methodology, al-

though many psychologists are conscious of the difficulty and troubled by it at a "semi-intuitive" level.

A large number of psychological studies use groups drawn from such selected sub-populations as college students, hospital inmates, nursery school children, etc. While the use of such sub-populations is perfectly proper, studies based on these groups have frequently been used to draw conclusions not restricted to the sub-population sampled. Such conclusions can be considered valid only when we know the facts regarding the selection of individuals present in the sub-population.

In the June 1946 *Biometrics Bulletin*, Berkson (4) discusses the limitations of fourfold table analysis with respect to data on the relationship of diabetes and cholecystitis in a hospital population. He assumes a general population in which the two conditions (diabetes and cholecystitis) are uncorrelated and points out that, if we assign to each condition a probability of a patient having the condition coming to the hospital for treatment, the hospital population of these two characteristics may show a correlation even though the two selective rates operate independently. Such a spurious correlation will exist in the hospital population unless the two selective rates are equal or one of the selective rates is zero (i.e. the probability of coming to the hospital is not affected by one of the conditions). Berkson notes:

Except for such cases there does not appear to be any ready way of correcting the spurious correlation existing in the hospital population by any device that does not involve the acquisition of data which would themselves answer the primary question. For instance the device sometimes used of setting up in the hospital sample a one-to-one control so that both groups examined have the same number of cases and are identical as regards say, age and sex, does not touch the difficulties referred to here.

Berkson's comments have much wider applications than hospital populations and fourfold tables. The problem he describes will exist with regard to any type of correlational analysis applied to any selected sub-population. The likelihood of selective bias is increased if the selective processes which generate the sub-population do not operate independently. Berkson points out:

The assumption that these probabilities operated independently in an individual who is suffering from more than one disease is doubtless oversimple. In general we may guess that, if a patient is suffering from two diseases, each disease is itself aggravated in its symptoms and more likely to be noted by the patient. So far as this difference of fact from assumption goes, its effect would be to increase relatively the representation of multiple diagnoses in the hospital, and in general to increase the discrepancy between hospital and parent population even more than if the probabilities were independent.

In a great many cases sub-populations arise from compound rather

than independent selective factors. College student groups are a selected population with reference to both intelligence and economic status and the selection is undoubtedly not independent for the two factors. Assume that a given college admits only U. S. born males in the age range 17 to 23 years, and therefore, that the population from which the college students are drawn may be defined as all native-born U.S. males aged 17 to 23 inclusive. Suppose further that the college admits 3% of those individuals in the population having I. Q.'s of 120 or more and 2% of those individuals coming from families with annual incomes of \$15,000 or more. If the admission process operated independently with respect to the two factors, we should expect that the college would admit 4.94% of the group of individuals having both an I. Q. of 120 or more and a family income of \$15,000 per year or more (i.e. 3% on the basis of I.Q. only, plus 2% of the remaining 97% on the basis of family income only). Actually the two factors may reinforce each other so that a larger number of individuals with both high I.Q. and high family income would apply for and be admitted to the college than would be expected from the consideration of the separate probabilities.

There are presented below selected results from two recently published research studies. The material presented does not represent the whole of either study. The present analysis covers only a small part of the conclusions of each study. We are concerned only with the particular phases of the studies presented herein and not with the validity of the studies as a whole. However, there are no facts presented in the complete published reports which would require modification of the analysis presented herein.

STUDY I: ALLPORT AND KRAMER, "SOME ROOTS OF PREJUDICE"

Allport and Kramer (1) studied prejudices of 437 college undergraduates enrolled in elementary courses in psychology at Dartmouth, Harvard and Radcliffe. They note:

While there is reason to suppose that college students (perhaps especially those studying psychology) are disposed to give less prejudiced responses than a more unselected population, still there is no ground for believing that their life-experiences with which the questionnaire deals, are in any significant way atypical.

As will be seen from the analysis presented below, the selection of the group studied by Allport and Kramer may very well influence the results obtained.

Allport and Kramer measured prejudice by use of a set of statements concerning Negroes, Jews and Catholics. These were scored on a 4-point scale in such manner that a high score indicated most prejudice

TABLE I
POPULATION AND SELECTION FACTORS WHICH COULD PRODUCE RESULTS
SHOWN BY ALLPORT AND KRAMER

	Hypothesis I*		Hypothesis II		Hypothesis III	
	Less Prejudiced	More Prejudiced	Less Prejudiced	More Prejudiced	Less Prejudiced	More Prejudiced
A. <i>Population Number in Each Group</i>						
Both parents college graduates	.65A	.35A	.40A	.60A	.40A	.60A
One parent college graduate	.65B	.35B	.40B	.60B	.50B	.50B
Neither parent college graduate	.65C	.35C	.40C	.60C	.65C	.35C
B. <i>Proportion of Population in Sample</i>						
Both parents college graduates	.03059	.03737	.04000	.01756	.04000	.01756
One parent college graduate	.01074	.01768	.02000	.01182	.01500	.01330
Neither parent college graduate	.00424	.01123	.01000	.00952	.00424	.01123
C. <i>Number in Sample</i>						
Both parents college graduates	.01988A	.01308A	.01600A	.01054A	.01600A	.01054A
One parent college graduate	.00698B	.00619B	.00800B	.00709B	.00750B	.00665B
Neither parent college graduate	.00275C	.00393C	.00400C	.00571C	.00275C	.00393C
D. <i>Sample Distribution</i>						
Both parents college graduates	60.3%	39.7%	60.3%	39.7%	60.3%	39.7%
One parent college graduate	53.0%	47.0%	53.0%	47.0%	53.0%	47.0%
Neither parent college graduate	41.2%	58.8%	41.2%	58.8%	41.2%	58.8%

* This hypothesis assumes independent selection rates of:

Less prejudiced
More prejudiced
Both parents college graduates
One parent college graduate
Neither parent college graduate

and a low score least prejudice. Table 21 of the report presents the following distribution:

PERCENTAGE DISTRIBUTION OF PREJUDICE SCORES AS
A FUNCTION OF PARENTAL EDUCATION

	<i>Less Prejudiced</i>	<i>More Prejudiced</i>
Both parents college graduates	60.3	39.7
One parent college graduate	53.0	47.0
Neither parent college graduate	41.2	58.8

In the above table "less prejudiced" denotes those subjects with scores below the median for the entire group of 437 undergraduates and "more prejudiced" denotes subjects with scores above the median. Allport and Kramer conclude from this table that

... students' prejudice scores bear an appreciable relation to the educational level of their parents. This finding is in line with the knowledge we now have that higher educational achievement on the part of an individual tends to be associated with fewer hostile attitudes. The novelty of the present finding consists in the demonstration that the favorable association between educational level and freedom from prejudice reaches back into the preceding generation: the more educated the *parent* the more free is the *child* from prejudice.

While the above finding may be quite reasonable, it is not the only interpretation possible of Allport and Kramer's results. Table I shows (as Part A) three hypothetical populations from which Allport and Kramer's sample might have been drawn. Selection ratios shown in Part B of Table I would yield from the hypothetical populations a distribution in the sample identical with that presented by Allport and Kramer in their Table 21.

Allport and Kramer's sample could have been drawn (as they assume implicitly) from a general population in which children of college graduates are less prejudiced than children of parents who are not college graduates. However, the sample could have come from populations with no difference in the degree of prejudice among the three groups (as in Hypotheses I and II of Table I) or from a population in which children of college graduates were more prejudiced than children of parents who are not college graduates (as in Hypothesis III of Table I).

No one can demonstrate that any one of the three hypotheses presented in Table I is correct or that the hypothesis that Allport and Kramer's group represented an unselected sampling of the general population is more or less likely than any of the three hypotheses presented in Table I. It is evident that no final conclusions are possible in the absence of information regarding the actual selection rates of the sample.

It should be noted that the three hypotheses presented in Table I are not unique. In the absence of knowledge about the actual selection rates, there are an infinite number of sets of possible populations and selection rates which could yield the results obtained by Allport and Kramer.

STUDY II: BERDIE, "RANGE OF INTEREST AND PSYCHOPATHOLOGIES"

Berdie (2) has developed a list of items to be used in measuring range of interests. Subjects check on a list of 22 activities, those which they "like to do." The 22 items include such activities as playing checkers, dancing, going on dates, hiking, reading, playing baseball, watching baseball, etc. Berdie administered this check list (orally to some groups and in printed form to others) to recruits at a U.S. Marine Corps recruit training station. He found that recruits recommended for inaptitude discharges (by the psychiatric unit) checked significantly fewer interests than did "normal" recruits.

In a second study (3), Berdie reports on range of interests for a group of 228 marine recruits recommended for inaptitude discharges. Berdie notes that the "abnormal" recruits contained "a heterogeneous collection of diagnoses, including various types of neuroses, constitutional states and organic conditions." For each diagnostic group, Berdie presents separate mean interest scores.

Questions can, of course, be raised regarding the adequacy of diagnoses and the meaning of the diagnostic categories used. This fact, in itself, operates as a "selective factor." Berdie states that the number of cases in each diagnostic group "can not be interpreted as showing how frequently these diagnoses occur in the Marine Corps or Naval Service, as no attempt was made to obtain a representative sample." It is obvious, of course, that even a representative sample of Marine Corps recruits with a given psychiatric diagnosis is not necessarily a representative sample of all individuals in the population having the given condition nor even of all individuals in the Marine Corps in a given psychiatric category (since undetected conditions must obviously have existed).

Berdie's group included 13 men classified as "epileptic" who showed a mean interest score of 10.23 items checked and 20 men classified as "constitutional psychopathic state-inadequate personality" who showed a mean interest score of 5.80. These two groups had respectively the highest and lowest mean interest scores of the 10 diagnostic categories reported in the study. Berdie states that "the probability of this difference" lies between the .01 and .02 levels.

In Table II are presented two hypothetical populations and sets of

TABLE II
POPULATIONS AND SELECTIVE FACTORS WHICH COULD
PRODUCE RESULTS SHOWN BY BERDIE

Interest Score	Population Distribution	A. HYPOTHESIS I CPS-Inadequate Personality		Epilepsy	
		Selection Rates	Distribution in Sample	Selection Rates	Distribution in Sample
0	.03479	.002868	.0998	.001185	.0412
1	.03715	.002624	.0975	.001179	.0438
2	.03930	.002392	.0940	.001171	.0460
3	.04122	.002172	.0895	.001159	.0478
4	.04291	.001960	.0841	.001145	.0491
5	.04438	.001760	.0781	.001129	.0501
6	.04563	.001568	.0716	.001110	.0507
7	.04663	.001392	.0649	.001090	.0508
8	.04743	.001224	.0580	.001069	.0507
9	.04800	.001064	.0511	.001046	.0502
10	.04834	.000920	.0445	.001023	.0495
11	.04844	.000784	.0380	.001000	.0484
12	.04834	.000660	.0319	.000977	.0472
13	.04800	.000544	.0261	.000954	.0458
14	.04743	.000444	.0210	.000931	.0442
15	.04663	.000348	.0162	.000910	.0424
16	.04563	.000268	.0122	.000890	.0406
17	.04438	.000200	.0089	.000872	.0387
18	.04291	.000140	.0060	.000855	.0367
19	.04122	.000088	.0036	.000841	.0347
20	.03930	.000052	.0020	.000829	.0326
21	.03715	.000024	.0009	.000821	.0305
22	.03479	.000008	.0003	.000815	.0284
Total	1.00		1.00		1.00
Mean	11.00		5.80		10.23
Variance	40.00		20.65		39.41
S. D.	6.32		4.54		6.28
B. HYPOTHESIS II					
0	.02	.002340	.0468	.000376	.0075
1	.03	.002121	.0636	.000383	.0115
2	.04	.001912	.0765	.000407	.0163
3	.05	.001715	.0858	.000446	.0223
4	.07	.001529	.1070	.000500	.0350
5	.08	.001355	.1084	.000571	.0457
6	.09	.001191	.1072	.000656	.0590
7	.10	.001039	.1039	.000758	.0758
8	.10	.000898	.0898	.000875	.0875
9	.09	.000769	.0692	.001007	.0906
10	.08	.000650	.0520	.001156	.0925
11	.06	.000543	.0326	.001319	.0791
12	.05	.000448	.0224	.001499	.0750
13	.04	.000363	.0145	.001694	.0678
14	.03	.000290	.0087	.001905	.0572
15	.02	.000228	.0046	.002131	.0426
16	.02	.000177	.0035	.002373	.0475
17	.01	.000138	.0014	.002630	.0263
18	.01	.000110	.0011	.002903	.0290
19	.01	.000093	.0009	.003192	.0319
Total	1.00		1.00		1.00
Mean	7.95		5.80		10.23
Variance	16.85		11.78		18.64
S. D.	4.10		3.43		4.32

selective sampling rates for individuals diagnosed as epileptic and as "CPS-inadequate personality." In both hypotheses it is assumed that the general populations for both psychiatric groups have identical interest score distributions but that a difference in selection rates for the sample exists. Such differences might arise from any of the selective factors noted above (i.e. selection from the general population for Marine Corps service, selection from Marine Corps inaptitude discharges for inclusion in the sample or faulty diagnosis). It might also arise from an association between range of interests and likelihood of detection of the condition, e.g. "constitutional psychopaths with inadequate personality" with a small range of interests might be more apt to come to psychiatric attention than those with a large interest range.

Whatever the true situation may be, we cannot conclude that differential sampling did not occur without specific data to demonstrate this fact. We cannot even assume that random selection was just as likely in this situation as differential selection. Such an assumption is the equivalent of the old fallacy of assuming that two events are equally probable in the absence of information regarding the actual probabilities (and then proceeding to demonstrate that there is almost certainly mammalian life on Mars).

From Table II it can be seen that the differences in mean score obtained by Berdie could have arisen through selective sampling. The result is not confined to Berdie's data. No confidence can be placed in obtained differences (regardless of "measures of significance") unless the experimenter has controlled carefully the method of sample selection and knows the probabilities of selecting a given element. If we have such knowledge, we can make proper estimates of population parameters and attach to our statistics properly determined measures of sampling error. Without such knowledge, generalization from sample to population is extremely risky. Since the application of a test of significance implies generalization from sample to population, all of our most valuable statistical techniques are meaningless without proper control of the *sampling method*.

It should be noted that a result which applies to sample means applies equally well to coefficients of correlation, analyses of variance, factor analysis, etc., since these techniques go back basically to differences in group means. It should be noted also that selective sampling affects the variance within groups as well as the variance between groups. In Table II the variance within the groups sampled is less than the population variance in three cases and greater in the fourth. The effect of selective sampling on total variance can only be determined

from a knowledge of the number in each group and the selective rates for each group. Thus, measures of correlation may be affected by selective sampling by distortion of both variance within groups and variance between groups.

SUMMARY

Through the use of two published psychological research studies, the possible effects of selective sampling are emphasized. The present analysis stresses the importance of control of the sampling process if valid generalizations are to be drawn from research data.

BIBLIOGRAPHY

1. ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *J. Psychol.*, 1946, 22, 9-39.
2. BERDIE, R. F. Range of interests. *J. appl. Psychol.*, 1945, 29, 268-281.
3. BERDIE, R. F. Range of interests and psychopathologies. *J. clin. Psychol.*, 1946, 2, 161-166.
4. BERKSON, J. Limitations of the applications of fourfold table analysis to hospital data. *Biometrics Bull.*, 1946, 2, 47-53.

ADDITIONAL TABLES OF THE PROBABILITY OF "RUNS" OF CORRECT RESPONSES IN LEARNING AND PROBLEM-SOLVING

DAVID A. GRANT

University of Wisconsin

In a previous paper (2), the writer demonstrated how "runs" of "correct" responses and grouping of "correct" and "incorrect" responses could be used as criteria of learning or problem-solution in experiments involving repeated trials.¹ The present paper has a double purpose: (a) to extend the tables of the probabilities of "runs" given in the earlier article; and (b) to discuss an important aspect of repeated statistical tests.

As pointed out in the original paper, a "run" of correct responses in a properly controlled learning or problem-solving experiment may permit the experimenter to conclude (at a specified confidence level, ϵ) that his subject has learned or solved the problem. A brief table of probabilities of chance runs was published in the original paper. Probabilities were given for various length runs of correct responses in total sequences of 16, 20, 25, and 50 trials, where the probability of a chance success on a single trial was 0.5. Because of the usefulness of the runs criteria, the limited table of the previous article has been extended in Table I of the present paper. In Table I, the probabilities of runs are given for total sequences of 16, 20, 25, 30, 35, 40, 45, and 50 trials where the probability of a chance success on a single trial is $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, or $\frac{1}{5}$.²

Two obvious trends will be noted in the probability entries of Table I, and both of these trends have direct implications applicable to the effective use of the runs criteria. First, the probability of any given run of correct responses decreases markedly as the probability (p) of a chance success on a single trial becomes small. Second, with any specified p , the probability of any given run of correct responses increases with an increase in the total number of trials in the series.

¹ Child (1) has subsequently pointed out an error in the original paper and showed how Yates' correction for continuity would, under certain conditions, improve the normal approximation formulae for the binomial.

² The entries of Tables I and II were computed for the writer by Marie H. Glissendorf, working under the supervision of Professor K. J. Arnold of the Mathematics Department, Director of the Computing Service of the University of Wisconsin. Each probability was computed from one or more of the formulae given by Uspensky (3, Ch. 3). Entries in Table I correct some errors in the fourth decimal place which occurred in Table I of the earlier paper (2, p. 278).

TABLE I

THE PROBABILITY OF A RUN OF AT LEAST S SUCCESSES IN N TRIALS WHEN THE PROBABILITY OF A SUCCESS IN A SINGLE TRIAL IS p .

$p = 1/2$								
$S \backslash N$	16	20	25	30	35	40	45	50
5	.1965							
6	.0929	.1223	.1578	.1918				
7	.0429	.0582	.0770	.0953	.1134	.1310	.1484	.1653
8	.0195	.0273	.0369	.0464	.0558	.0652	.0744	.0836
9	.0088	.0127	.0175	.0224	.0272	.0320	.0367	.0415
10	.0039	.0059	.0083	.0107	.0132	.0156	.0180	.0204
11	.0017	.0027	.0039	.0051	.0063	.0076	.0088	.0100
12	.0007	.0012	.0018	.0024	.0031	.0037	.0043	.0049
13	.0003	.0005	.0009	.0012	.0015	.0018	.0020	.0024
14	.0001	.0002	.0004	.0005	.0007	.0009	.0010	.0012
15	.0000	.0001	.0002	.0003	.0003	.0004	.0005	.0006
$p = 1/3$								
4	.1084	.1384	.1745	.2090	.2421	.2739		
5	.0341	.0448	.0580	.0711	.0839	.0966	.1091	.1214
6	.0105	.0141	.0187	.0232	.0277	.0321	.0366	.0410
7	.0032	.0044	.0059	.0075	.0090	.0105	.0120	.0135
8	.0010	.0014	.0019	.0024	.0029	.0034	.0039	.0044
9	.0003	.0004	.0006	.0008	.0009	.0011	.0013	.0014
10	.0001	.0001	.0002	.0002	.0003	.0004	.0004	.0005
$p = 1/4$								
3	.1601	.2002	.2476	.2922	.3342			
4	.0387	.0501	.0641	.0779	.0915	.1048	.1180	.1310
5	.0090	.0119	.0156	.0192	.0228	.0264	.0299	.0335
6	.0021	.0028	.0037	.0046	.0055	.0065	.0074	.0083
7	.0005	.0007	.0009	.0011	.0013	.0016	.0018	.0020
8	.0001	.0002	.0002	.0003	.0003	.0004	.0004	.0005
$p = 1/5$								
2	.4107							
3	.0889	.1124	.1410	.1687	.1955	.2214	.2465	.2712
4	.0169	.0219	.0282	.0345	.0407	.0468	.0529	.0590
5	.0031	.0042	.0054	.0067	.0080	.0092	.0105	.0118
6	.0006	.0008	.0010	.0013	.0015	.0018	.0021	.0023
7	.0001	.0001	.0002	.0002	.0003	.0004	.0004	.0005

The first feature mentioned above becomes important when it is desirable to minimize over-learning in experiments in which the subject is likely to show sudden "insightful" learning or problem-solution. In

TABLE II

THE LARGEST NUMBER OF TRIALS FOR WHICH THE PROBABILITY OF A RUN OF AT LEAST S SUCCESSES IS LESS THAN ϵ WHEN THE PROBABILITY OF A SUCCESS ON A SINGLE TRIAL IS p .

$p=1/2$						$p=1/3$					
S	ϵ					S	ϵ				
	.001	.005	.010	.050	.100		.001	.005	.010	.050	.100
4	—	—	—	—	5	4	—	—	—	8	14
5	—	—	—	6	9	5	—	5	7	21	41
6	—	—	—	10	16	6	—	9	15	60	119
7	—	—	7	17	31	7	8	21	38	173	350
8	—	8	11	31	59	8	16	55	105	510	1042
9	—	12	17	59	113	9	37	155	304	1521	3117
10	10	18	28	112	222	10	97	452	898	4551	9340
11	13	29	50	218	439						
12	18	51	92	429	871						
13	27	93	175	850	1735						
14	44	176	341	1692	3463						
15	78	341	671	3374	6917						

$p=1/4$						$p=1/5$					
S	ϵ					S	ϵ				
	.001	.005	.010	.050	.100		.001	.005	.010	.050	.100
3	—	—	—	5	10	2	—	—	—	2	3
4	—	4	6	19	38	3	—	—	3	9	17
5	5	10	17	73	146	4	—	6	10	42	84
6	10	32	59	284	579	5	7	23	42	203	414
7	27	115	225	1125	2306	6	24	102	200	1006	2061
8	94	444	884	4486	9208	7	103	495	987	5014	10294

such cases, a reduction of the probability of a chance success on a single trial³ will reduce the length of the run necessary to indicate learning. Short runs might also be of value in detecting "hypothesis" behavior and other theoretically consistent but evanescent forms of behavior.

³ Reduction of the a priori probability of a chance success may be accomplished by adding to the number of possible but incorrect responses.

The second feature mentioned above leads to a difficulty when training is to be given until the subject's performance meets any constant criterion. Suppose, for example, an experimenter is running 25 trials a day on a simple discrimination problem in which the subject has to pick the correct *one* of *two* objects. He notes that $p = \frac{1}{2}$, and from Table I, that a run of *eight* or more correct responses ($P = .0369$) would be significant at the 5% level and decides to use this as his criterion of learning. Suppose that his subject fails to achieve this criterion the first day. The experimenter might then decide to continue training until, on some day, the criterion is met. With sufficient persistence on the part of the experimenter, the probability of achieving this criterion will approach 1.00 even though the subject always responds randomly with respect to the critical cue, throughout the experiment. Correct use of the runs criteria requires that all trials within a previously selected limited series be taken into account. This condition may be met in any of a number of ways. For example, the experimenter may decide *in advance* to limit his experiment to a given number of trials, say, 500. Or, again, he may decide *in advance* that the 25 trials shall be *the* one test series. Some limit must be set—in other words, *failure of the experiment as well as its success must be made possible*.

To make convenient the application of the runs criteria when a large total number of trials is contemplated, Table II is presented in which the maximum series length is given such that a run of at least S correct responses is significant at the given confidence level, ϵ .

In using Table II, the experimenter described above would presumably decide, in advance, that if the subject has not achieved the criterion in a certain number of trials, say, 429, the experiment will be terminated. If 429 were selected, the significant ($\epsilon < .05$) run would be 12 correct responses. Here the daily records would be pooled so that a run of correct responses might extend over two days or two 25-trial series. Again it will be noted, that had the single trial probability of a chance success been smaller than $\frac{1}{2}$, the significant run would have been quite short.

BIBLIOGRAPHY

1. CHILD, I. L. A note on Grant's "New statistical criteria for learning and problem solution." *Psychol. Bull.*, 1946, 43, 558-561.
2. GRANT, D. A. New statistical criteria for learning and problem solution in experiments involving repeated trials. *Psychol. Bull.*, 1946, 43, 272-282.
3. USPENSKY, J. B. *Introduction to mathematical probability*. New York: McGraw-Hill, 1937.

ENCYCLOPEDIA OF PSYCHOLOGY

A SPECIAL REVIEW*

S. RAINS WALLACE

Tulane University

This volume is an attempt to present the field of psychology in encyclopedic form. It consists of approximately 100 articles representing 82 contributors. Some of the articles discuss such broad subjects as *Child Psychology*, *Emotion*, *Personality*, or *Motivation*. Others—for example, *Aspects of Blindness*, *Psychological Psychodrama* and *National Council of Women Psychologists*—are confined to relatively narrow and specialized topics. In general, emphasis is placed upon contemporary problems, approaches, and findings, although an occasional article presents a somewhat exhaustive historical account. Some contributors attempt to give a thorough review of their topics with carefully annotated references to the literature. Many writers content themselves with general discussions in which their own interests and theories take precedence over the work of others. From the viewpoint of both subject matter and style, therefore, there is an inexcusable lack of uniformity or consistency.

In a preface, the editor writes that one purpose of this volume was "... to emphasize some of the trends in contemporary psychology which seem to have supplanted much of the traditional material." This appears to indicate that some general plan was followed in determining the apportionment of space to the various subjects. Unfortunately, such a plan is not readily discernible. Certainly one hesitates to believe that the allotment of more space to *Parapsychology* and *Pseudopsychology* than to *Learning*, *Experimental Psychology*, *Industrial Psychology* or *Aptitudes and Aptitude Testing* is indicative of the Editor's opinions about current trends in the field. While some plan may be discovered in the fact that topics which relate to abnormal psychology and psychiatry are assigned over 200 pages of the 897 total, one must wonder if it were really planned that *Psychosomatics* should receive 43 pages, *Psychotherapy*, 16, and *Religion and Psychiatry*, 17. Similarly it is difficult to grasp a plan which leads to the inclusion of separate sections on *Frames of Reference in Psychology*, *Phenomenology*, *Philosophical Psychology*, *Points of View*, *Scientific Law in Psychology*, and *Systematic Psychology*.

* HARRIMAN, PHILIP L. (ED.) *Encyclopedia of psychology*. New York: The Philosophical Library, 1946. Pp. vii+897.

Equally difficult is the explanation of a 13-page section on *Language and Psychology*, a 9-page section on *Language, Psychology of*, and a 33-page section on *Semantics, Language*. When these assignments are compared to the 11 pages given to *Learning*, the 4 pages given to *Industrial Psychology*, and the 16 pages given to *Personnel, Teacher*, the existence of any plan at all is placed in doubt.

This doubt is not dissipated by a closer examination of the book. Instead, as in "Alice in Wonderland," things get "curiouser and curiouser." It is a disturbing discovery that 8 pages are devoted to a discussion of the *Eye*, including some material which is largely duplicated in a section on *Color*, but that only 4 pages are given to an anatomical description of the ear, allowing a few paragraphs in the section on *Receptors and Effectors* for the facts of audition. Does this mean that current trends in psychology have forced hearing out of style? Is there significance in the fact that Jung is the only psychologist to appear in a section under his own name or that he, Adler, Porteus, and Rorschach are the only individuals referred to in the index?

Elsewhere there is further evidence that the editor failed to plan or, indeed, to edit. According to the Preface, the volume was intended "... to meet the requirements of the serious investigator who wishes to acquaint himself with various topics in modern psychology which lie outside his field of special interest and competence." It was also supposed to be serviceable "in the enrichment of various topics by initiating the student into an intensive exploration of the primary data." How will the serious investigator react when he finds that some articles are accompanied by complete bibliographies while others have none? It is to be hoped that he will not suppose that bibliographies are omitted only for those topics for which publications are lacking. Certainly, it would be unfortunate if a student arrived at that conclusion when he discovered his intensive exploration of the primary data in the field of learning blocked by the absence of a bibliography while one of 37 titles beckoned him into the field of *Porteus Maze Tests*.

Other evidence of inadequate editing is found in the fact that the bibliographies are presented in inconsistent formats which appear to depend upon the habits and vagaries of the individual contributors. This defect is only one of a number of symptoms. For example, the section on *Genetic Psychology, Experimental*, refers by name and number to over 180 sources but the editor somehow failed to append the bibliography. These faults do not increase the reader's confidence. They do increase the sympathy he feels for the contributor who, in her article on *Color*, blithely wrote "See heredity," and who now knows that there

is no section on heredity. The "index" of this encyclopedia is a 3-page table of contents with an attempt at cross-indexing which is exemplified by the entry, "Ear—see Eye; Receptors and Effectors." Incidentally, the "index" includes no page numbers.

Some may feel that these criticisms are of a petty variety. They are presented, however, to support a grave charge. In brief, it is believed that the contents of this volume were amassed by the process of selecting certain contributors, and in general, of allowing their own interests, freedom of time, notions of style, etc., to determine what, how, and how much they wrote. It is submitted that even if contributors are maximally expert in their particular fields, a useful reference volume cannot be produced unless someone assumes responsibility for over-all planning and presentation. In this case we may question not only the discharge of this responsibility but also the assertion of the Editor that these articles are written "by those who have achieved distinction through contributions to special areas within the science of psychology." Many of the contributors are indeed eminent in the field of which they write. But many others are relatively unknown. Whether the writers deserve to be unknown is not to be discussed here. We may, however, leave it to the reader to reconcile the editor's statement with the fact that the section on *Emotion, Frames of Reference in Psychology, Jung, Personal Document Analysis*, and *Phenomenology* are the work of a single contributor.

Admittedly, this is not a pleasant review. It is not intended to be. The title of this book is such as to place it almost inevitably in every college library and in the hands of many laymen who honestly desire to extend their knowledge of the field. At its best it will give students and laymen alike a mistaken notion about the nature of psychology as a science. At its worst, it will make them think they know more than they do or will make the more perspicacious of them think psychologists agree even less than they do. The thinking student who reads the section on *Conditioning* and is told that "... Hull has formulated by far the most extensive and consistent theory of learning thus far offered ..." is going to perceive that something is radically wrong when he reads the section on *Learning* and finds that Hull's work is not mentioned. Something is surely wrong; but with the book, not with the field. Should not something be done to protect psychology from this sort of thing?

THE RELATION OF EMOTIONAL ADJUSTMENT TO INTELLECTUAL FUNCTION—A NOTE

ROBERT E. HARRIS AND CLARE WRIGHT THOMPSON

*University of California Medical School, Department of Psychiatry and
The Langley Porter Clinic*

It has often been asserted that changes in a child's emotional adjustment are reflected in corresponding changes in his tested intelligence. Most clinical psychologists have had occasion to report that a measured IQ has seemed too low an estimate of a disturbed or anxious child's "real" intelligence. This reservation is usually based on observations of anxiety and non-intellectual attitudes which operate to decrease efficiency in the test situation. The statement implies that if these factors were not present the measured IQ would be higher. Since tests of intellectual ability are usually standardized on non-clinic populations in whom these factors are minimal, it is sometimes necessary to make such qualifications.

However, it is commonly asserted that there is a more pervasive kind of impairment of intellectual function which operates to lower efficiency, not only in the test situation but in other situations demanding intellectual effort as well. It is further asserted that changes in the life situation operating over the course of time, or expedited by psychotherapy or other manipulation, so modify these inhibiting factors that subsequent IQ's are significantly different from the initial measure. Demonstration of the existence of such factors requires, of course, measurement over a period of time of changes which are greater than might be expected from the inadequacies of the test and the difficulties inherent in testing emotionally disturbed children.

The most recent study purporting to demonstrate such changes is that of Despert and Pierce (1). Their purpose was to "ascertain whether any relation can be established between emotional adjustment as shown in the total record and intellectual function as determined through psychometric testing" (p. 7). The revised Stanford-Binet was administered to 39 children before and after attendance at nursery school. Available also was information from which they made evaluations of personality, of socio-emotional adjustment and of the family situation. After the administration of the final test, the records were examined for factors which might account for observed increases or decreases in IQ. According to the authors: "The results of the psychometric tests and the findings related to socio-emotional adjustment were analyzed *inde-*

is no section on heredity. The "index" of this encyclopedia is a 3-page table of contents with an attempt at cross-indexing which is exemplified by the entry, "Ear—see Eye; Receptors and Effectors." Incidentally, the "index" includes no page numbers.

Some may feel that these criticisms are of a petty variety. They are presented, however, to support a grave charge. In brief, it is believed that the contents of this volume were amassed by the process of selecting certain contributors, and in general, of allowing their own interests, freedom of time, notions of style, etc., to determine what, how, and how much they wrote. It is submitted that even if contributors are maximally expert in their particular fields, a useful reference volume cannot be produced unless someone assumes responsibility for over-all planning and presentation. In this case we may question not only the discharge of this responsibility but also the assertion of the Editor that these articles are written "by those who have achieved distinction through contributions to special areas within the science of psychology." Many of the contributors are indeed eminent in the field of which they write. But many others are relatively unknown. Whether the writers deserve to be unknown is not to be discussed here. We may, however, leave it to the reader to reconcile the editor's statement with the fact that the section on *Emotion, Frames of Reference in Psychology, Jung, Personal Document Analysis, and Phenomenology* are the work of a single contributor.

Admittedly, this is not a pleasant review. It is not intended to be. The title of this book is such as to place it almost inevitably in every college library and in the hands of many laymen who honestly desire to extend their knowledge of the field. At its best it will give students and laymen alike a mistaken notion about the nature of psychology as a science. At its worst, it will make them think they know more than they do or will make the more perspicacious of them think psychologists agree even less than they do. The thinking student who reads the section on *Conditioning* and is told that "... Hull has formulated by far the most extensive and consistent theory of learning thus far offered ..." is going to perceive that something is radically wrong when he reads the section on *Learning* and finds that Hull's work is not mentioned. Something is surely wrong; but with the book, not with the field. Should not something be done to protect psychology from this sort of thing?

THE RELATION OF EMOTIONAL ADJUSTMENT TO INTELLECTUAL FUNCTION—A NOTE

ROBERT E. HARRIS AND CLARE WRIGHT THOMPSON

*University of California Medical School, Department of Psychiatry and
The Langley Porter Clinic*

It has often been asserted that changes in a child's emotional adjustment are reflected in corresponding changes in his tested intelligence. Most clinical psychologists have had occasion to report that a measured IQ has seemed too low an estimate of a disturbed or anxious child's "real" intelligence. This reservation is usually based on observations of anxiety and non-intellectual attitudes which operate to decrease efficiency in the test situation. The statement implies that if these factors were not present the measured IQ would be higher. Since tests of intellectual ability are usually standardized on non-clinic populations in whom these factors are minimal, it is sometimes necessary to make such qualifications.

However, it is commonly asserted that there is a more pervasive kind of impairment of intellectual function which operates to lower efficiency, not only in the test situation but in other situations demanding intellectual effort as well. It is further asserted that changes in the life situation operating over the course of time, or expedited by psychotherapy or other manipulation, so modify these inhibiting factors that subsequent IQ's are significantly different from the initial measure. Demonstration of the existence of such factors requires, of course, measurement over a period of time of changes which are greater than might be expected from the inadequacies of the test and the difficulties inherent in testing emotionally disturbed children.

The most recent study purporting to demonstrate such changes is that of Despert and Pierce (1). Their purpose was to "ascertain whether any relation can be established between emotional adjustment as shown in the total record and intellectual function as determined through psychometric testing" (p. 7). The revised Stanford-Binet was administered to 39 children before and after attendance at nursery school. Available also was information from which they made evaluations of personality, of socio-emotional adjustment and of the family situation. After the administration of the final test, the records were examined for factors which might account for observed increases or decreases in IQ. According to the authors: "The results of the psychometric tests and the findings related to socio-emotional adjustment were analyzed *inde-*

pendently" (italics theirs). However, they do not use the word "independently" in its usual sense; apparently they did not attempt to predict the measured intellectual changes from personality data alone. Rather, the procedure seems to have been to examine the test-retest differences, to select those cases showing greatest change and then to examine the other data for possible reasons. In every case showing decline it was possible to find some plausible reason such as parental discord, the birth of a sib, a hostile nurse, etc. The factors associated with rises were more obscure, the phrase, "improved emotional and social adjustment," appearing several times. The authors do not hesitate to ascribe causal effect to these factors.

However, examination and recalculation of their data indicate that the changes can be explained by a simpler hypothesis. It is possible, knowing only the initial IQ's and the standardization data of the test, to predict for the group as a whole the mean IQ and its variability on a second test. It is possible also to predict for any given IQ range within the group both the direction and the magnitude of change on a second test. These predictions agree almost perfectly with the results presented by Despert and Pierce. For example, there are two cases with IQ's in the range above 150; one child tested at 155, another at 157. On statistical grounds alone (see below), one would predict that on a retest (an alternative form administered the same day) initial IQ's of 155 would decline to 142, and initial IQ's of 157 to 143. The actual findings are 141 in the first case and 144 in the second. These two are among the cases for whom the authors produce data accumulated over a period of time to account for a "significant" change (10 points or more). Also on statistical grounds, one is not surprised to find that those cases showing significant declines averaged 34 points higher initially than a group showing comparable rises.

The initial mean IQ of the ten cases selected by Despert and Pierce as showing a significant change downward is 140 and that of the group changing upward 106. The uncorrected correlation ratio between initial score and loss is .80 ($N=39$, $P<.001$). Apparently selecting for direction and size of change has isolated groups initially quite different from each other. However, on the second test, these groups approach each other very closely; the group showing losses averaged 127 and the group showing gains, 121. Thus both groups tend to approximate the mean of all cases. A conclusion one might draw is that if a child scores high initially, then bad things are going to happen to him, and if he scores low, good things are going to happen so that he ends up closer to the mean of the group in either case. A more credible explanation is in terms of statistical regression.

Statistical regression is the name for the phenomenon that a group of scores selected because of their deviation from the mean of a population will, on a second test, more closely approximate the population mean. A score on a test may be thought of as having two components: the "true" amount of the trait measured plus a variable error. If from a group of test scores, one selects the highest, the effect is to increase the probability of selecting those which have associated with the true score the greatest amount of plus error. On a second test, the errors would be just as likely to be in the opposite direction and consequently the second set of scores would be lower. The converse would be true for scores below the mean. This principle is nicely illustrated by the Despert-Pierce data. It is possible to predict precisely the amount by which any selected group of scores will regress toward the mean. Since the amount of regression varies inversely with the reliability of the test, it is necessary to know the reliability coefficient.¹ The only other data needed are the initial mean and standard deviation for the total group, and, if they change significantly, these same values for the second administration.

In the Despert-Pierce data the means from the two administrations were within one point of each other. The standard deviations varied by the amount and in the direction to be expected for children of this age (3). Predictions of regressed mean values were made for various IQ levels from the usual regression equations. If the total group is divided at the mean, those scoring above the mean averaged 135 on the first test. The predicted mean for the second test was 130, exactly the same as the obtained figure. For the cases scoring below the mean the average score was 105 on the original test and 114 on the second test. The predicted mean was 112. This is an insignificant difference (P equals 0.80 for 17 d. f.). Breaking down the groups more finely gives the data shown in Table I.

Examination of this table makes it clear that those levels farthest from the mean regressed the most. With this small number of cases, it is impressive how closely the predicted values approach the observed. Obviously no hypothesis other than regression is necessary to account for the changes. To isolate particular individuals from these groups as

¹ The test-retest correlation (ρ) calculated from the Despert-Pierce data is .69. Because of the time interval between the tests and the (probably) different functions measured at the two age levels, this figure cannot be accepted as a measure of reliability. In the calculations to follow we have used the reliability coefficients appropriate to age and IQ levels as reported by McNemar (3). When we made the same calculations using the value .69, the observed and predicted values were even more similar than those reported in the text.

showing spectacular changes is to select cases where the errors have been in opposite directions on the two administrations, and to assign psychological significance to statistical artifact.

We have labored this criticism so far because of the importance of the problem which Despert and Pierce purported to study. That what are called socio-emotional factors interfere with (or facilitate) intellectual achievement there can be no doubt. As an example one might cite the nine percent of men in the Stanford Gifted Group who as young

TABLE I
COMPARISON OF ACTUAL RETEST MEANS AND THOSE
PREDICTED FROM REGRESSION EQUATIONS

IQ Range*	N	Mean of first test	Mean of second test		Dis- crepancy
			Predicted	Observed	
67-103	7	93.0	104.4	106.3	+1.9
104-121	11	113.0	117.2	118.9	+1.7
122-139	15	129.7	127.4	125.8	-1.6
140-157	6	147.7	137.8	139.2	+1.4

* Class intervals based on size of standard deviation.

adults failed to rise above the level of semi-skilled labor (4). Similarly every one knows examples of superior achievement in individuals of mediocre ability. But that the IQ itself is subject to socio-emotional factors is not obvious nor, as far as we know, has it been demonstrated. Proof would require an experimental design less contaminated and statistically more sophisticated than the study discussed here. Our recalculation of the data offers strong presumptive evidence that the influence of such factors is minimal.

There remains the possibility that more severe forms of maladjustment than are represented in a nursery school population may depress measured IQ's and that more intensive therapy may raise them. Published data on this point are equivocal at best (2). Our own data, based on 30 clinic children, are too unwieldy for a simple quantitative summary. Study of individual cases, taking account of age and IQ level (factors related to the size of the standard error), special handicaps, conditions of testing, etc. has led to the tentative generalization that within wide ranges of different kinds of personality and behavior problems,² and within a wide range of "cooperation" in the test situation, a

² We do not mean to include here psychotic conditions, or organic brain disease or other physical conditions known to affect cortical function.

measured IQ remains fairly stable, even after intensive therapy and improvement in adjustment. The fact that changes in IQ have been so small has often been disturbing to therapists working with the patients. A number of factors are at work here. First there is the fact that in our culture a high IQ is a "good thing." Second, a therapist has some personal stake in the child and his improvement—by identification with the child, as a validation of his therapeutic skill, etc. Third, because there are no unambiguous indices of improvement in adjustment, the IQ is frequently seized upon, erroneously, of course, as such a measure. This is an unfortunate state of affairs, particularly because not only psychotherapists, but also educators, workers in social agencies, and others are motivated to look eagerly for rises in IQ.

A further implication of the principle of regression for clinical work should be mentioned. Schools for mental defectives and other agencies concerned with children of below average intelligence often have the experience of obtaining higher IQ's on re-testing than those reported by a referring agency. This is a consequence of the fact that a low IQ is one of the factors for which children are referred to such agencies. If the principle of regression is not understood, there are at least two possible misinterpretations of the higher IQ's on re-test. If the second test is given shortly after the first, doubt may be cast on the adequacy of the first testing. If the second test is delayed until after a period of training or therapy, the conclusion may be drawn that the program has effected the rise in IQ. As in the case of the Despert-Pierce data, the effect of regression is to move the scores closer to the mean of the group from which they were selected, in this case, the mean of the general population.

It is easy to understand the motivation of those who would like to raise IQ's. It should be obvious, however, that improved socio-emotional adjustment is an end in itself, and any program effecting it need not depend on changes in measured intelligence for confirmation of its success.

BIBLIOGRAPHY

1. DESPERT, J. LOUISE, & PIERCE, HELEN O. The relation of emotional adjustment to intellectual function. *Genet. Psychol. Monogr.*, 1946, **34**, 3-56.
2. DULSKY, S. G. Affect and intellect: an experimental study. *J. gen. Psychol.*, 1942, **27**, 199-220.
3. McNEMAR, Q. *The revision of the Stanford-Binet scale*. Boston: Houghton Mifflin, 1942.
4. TERMAN, L. M., & ODEN, MELITA. Correlates of adult achievement in the California gifted group. In *Intelligence: its nature and nurture*, Thirty-Ninth Yearbook of the National Society for the Study of Education, Part I, 74-89. Bloomington, Ill.: Public School Publishing Co., 1940.

CORRIGENDA

In re-writing and re-copying the material for my article entitled "The Detection and Treatment of Accident-Prone Drivers," which appeared in the *Psychological Bulletin* in November, 1946 (vol. 43, pp. 489-532) I overlooked several small errors which may annoy the reader unless he correct them. They are as follows:

- p. 492, 2d line from bottom of text:
For -0.21 put -0.021
- p. 492, second footnote:
For -0.27 put -0.027
For -0.26 put -0.026
For 0.21 put 0.021
For "counter" put "according."
- p. 493, line 3:
For 0.37 put 0.037
For 0.18 put 0.018
- p. 494, last line:
For 0.84 put 0.79
- p. 500, line 2:
For 19 put 10
- p. 510, par. 1, line 3:
For r_{12} put r^2_{12}

H. M. JOHNSON
Tulane University

BOOK REVIEWS

DUNLAP, KNIGHT. *Religion: its functions in human life*. New York: McGraw-Hill, 1946. Pp. xi+362.

This book is "a study of religion from the viewpoint of psychology," an attempt to bring together in systematic fashion the characteristic phenomena of religious behavior and on the basis of these data to formulate a theory of religion which explains them. The theory thus arrived at is stated as follows:

... in the early stages of religious development religion can be described as the attempt to do what man does not know how to do and to know the unknowable, ... in its present stage ... religion is the institution or feature of culture, which undertakes, in the service of mankind, the functions for which no other institution is as yet adequately prepared (p. 321).

The book is basically a volume in comparative religion. Its plan consists in a series of religious topics, under each of which is presented typical comparative religious data followed by observations concerning the psychological significance of these data. These topics include: religion as morality, concepts of divinities, the concept of sin, the concepts of the other world, the evolution of divinities, the role of desire in religion, religious symbolism, conversion, the nature of religion, and the future of religion in civilization, to mention only a few.

The author makes generalizations repeatedly in the comparative religion portions of the book which can be made only by authorities in the field. In the psychological sections of the book, several hundred (estimated) generalizations are made, of which his over-all description of religion, previously quoted, is the central one. Many of these generalizations seem to be justified by the data on which they are based, assuming the completeness of those data. For example there are many such well substantiated observations as these: "Gods are not essential for religion ... All religions, however, ... reach a stage in which gods and goddesses appear and multiply (p. 98). ... only the concept of a personal god can satisfy man's feeling toward divinity. ... The substitution of an impersonal world power for a personal god is a regression toward a more primitive conception" ... (p. 336). Many of these conclusions however would seem to be easily replaced by other and often contradictory hypotheses to explain the same data.

There is nothing in the volume that remotely touches on experimental psychology. The major concepts in the psychology of personality are only occasionally mentioned and then not systematically. Many psychologists will be astonished at this statement: "We make an unjustified assumption even if we ascribe motivation to animals, for motivation involves desire, which is a complex process probably above

the mental level even of the ape" (p. 126). This is especially interesting because it is followed by a list of primary desires including the desire for food and drink, rest, activity, and the erotic desire, which do not seem to be absent from most animals.

Three religious concepts are dealt with in generalizations which appear to be less profound than the problems themselves deserve. They are the concept of salvation, the problem of evil, and the problem of immortality. Two quotations are illustrative. "'Salvation' may be loosely defined as the process, or the condition, of being saved; that is, of being rescued from, or protected from, some harm, danger, risk or hazard" (p. 82). This is purely negative. In explaining evil, Dunlap says: "God, creating man in His image, made him a person and thus permitted man to create evil. This may not be a satisfactory solution . . . but it is the best any theologian has been able to think out" (p. 254).

Those psychologists who seem to believe that an adequate interpretation of human personality and social behavior can be made without a serious evaluation of religious behavior ought to read this volume. However little or much one may agree with the author, he could not fail to be convinced of the importance of the field.

Those whose religious convictions make it difficult for them to study it with truly scientific objectivity will do well to examine this largely dispassionate discussion of religious ideas. It will show them other points of view than their own.

Despite this somewhat critical review, the reviewer finds himself in awe of and with considerable admiration for the courage of the author in undertaking so gigantic a task. The book will probably not receive nearly the breadth of reading it deserves. As a source book in religion and social psychology it must be given an important place in every library in which students investigate related problems.

ERNEST M. LIGON.

Union College.

FRIEDMAN, BERTHA B. *Foundations of the measurement of values: The methodology of location and quantification*. New York: Teachers College, Columbia University, 1946. Pp. viii+227.

Critical of the majority of present attitude and interest tests because "little attention was paid to the theories upon which the tests rested, or to the means used to validate the test," the author attempts to erect "a framework upon which value-testing instruments could be built."

For this purpose she examines, at times in exhaustive detail, a number of questions directly or indirectly related to the measurement of values. Among these are the definition, location, and quantification

of values, and types of scales which have been, or might be used in the measurement of values.

Since the purpose of value-measuring instruments is to aid in the prediction of conduct, values are defined in terms of behavior. As indices of value, it is proposed to use the way in which a person spends his time, money, or energy. Data may be obtained through overt behavior or through verbal statements. Each method has its inadequacies. There is a further problem in the naming and classification of values, but here, as elsewhere in measurement, decisions should take account of the purpose of the investigator.

Quantification of values may be secured through measuring the amount of time or money expended or the number of instances where X is chosen in preference to Y . However, the primary problem of quantification is the determination of the empirical meaning of scale scores and not the measurement of scale-point intervals. Psychological, and not merely logical, continua must be established.

Following the discussion of what constitutes measurement in the field of values—to which the greater portion of the book is devoted—the writer describes three experimental investigations of her own designed to contribute data on certain questions which she has previously raised.

The book contains a comprehensive analysis of many problems basic to the measurement of values. Pitfalls and difficulties of great variety are pointed out. In fact, the writer is led at one point to inquire, "Is the possibility of locating values, then, hopeless?" In reply she states, "But this inquiry, it must be remembered, did not begin with any assumption that values could always be accurately located, let alone accurately quantified. Instead, it set as its goal the improvement of the location and quantification of values, which it recognized was an activity that was being pursued almost wholly without concern for methods which would bring increasingly better results."

To many readers the chief interest of the book will lie in the specific questions it raises in regard to the Thurstone scales and other widely used scales for the measurement of attitudes or values.

While the author has undoubtedly made a contribution to the field of value-measurement, the discussion at times seems unnecessarily circuitous, and the reader is led to wonder whether the longest way round is really the shortest way home. And it must be added in all sadness that the exhaustive consideration of theory does not appear to the reviewer to have improved the "practice" in values measurement as represented in the techniques used by the author in her own experimental investigations.

ELIZABETH DUFFY.

The Woman's College of the University of North Carolina.

MAYS, A. B. *The concept of vocational education in the thinking of the general educator, 1845 to 1945*. University of Illinois Bulletin, 1946. Pp. 107.

This development of general educators' ideas provides background for understanding educational practices. It has wider significance than an historical account of slow progress from cultural to balanced-general-vocational philosophies of education. One may infer that during the last century two opposing forces controlled ideas: the traditional, advocating a cultural education for higher conformities to the regional culture; the newer, advocating vocational education for higher efficiencies to serve the needs of agriculture, commerce and industry. The general educators' ideas guarded the cultural features with few exceptions. "Freeing" ideas came from insights related to widespread survival needs and satisfaction drives of economic interests and local communities—a few becoming social movements, chiefly with Federal support.

Maladaptive social implications such as support for hardening unequal caste distinctions are seen for the cultural trend. Such insights are lacking for the economic pressures and competitive struggles and techniques, though they are implicit in symptoms of tensions and disturbances of thinking.

The author faces the future with optimism on the basis of the balanced solution, though this looks like an unstable affair. Progress for the atomic age should develop the abandonment of competitive struggles and wider educational diets than balancing political and economic interests alone. The author and his selected quotations neglect needs related to building a world culture to secure cooperation and security; and to implementing the drives of free individuals and groups to realize their personal, social, and creative interests and develop their distinctive talents.

JAMES L. GRAHAM.

Lehigh University.

KINGSLEY, HOWARD L. *The nature and conditions of learning*. New York: Prentice-Hall, 1946. Pp. xvi+579.

When one attempts to evaluate a new textbook two points of view compete for his attention as frames of reference. The first appraises the newcomer with reference to earlier texts and the trends of development which they exhibited. The second derives from experience with the needs which such a text must meet. Accordingly, on the one hand, the reviewer looks for changes in scope and emphasis in comparison with previous works in the field. Once these have been identified, he evaluates the salient points of the new work with respect to degree of approximation to the structure of what the reviewer regards as a desirable textbook.

Professor Kingsley's text, when viewed within these frames of reference, impresses the reviewer favorably on both counts. More than

any other text with which he is familiar, the book presents psychological material in a manner most likely to assist the student, preparing for an educational career, in acquiring the elementary background of principle and technique prerequisite to advanced study. According to the author, the treatment of the field of educational subjects:

... rests on the conviction that learning is the core of educational psychology, that laboratory research provides the most reliable data concerning the fundamental nature and conditions of learning, and that the educational psychology of learning must be based solidly on the fruits of the laboratory. . . . An effort has been made to secure a balance of experimental evidence, illustration, interpretation, summary, and application. Illustrations are provided from everyday life and from the classroom to promote the comprehension of principles and to help bridge the gap between the research laboratory and the classroom.

These quotations from the preface are amply substantiated in the body of the text. Part I deals with the development of the subject matter and techniques of educational psychology and their role in the educative process, as well as treating comprehensively the fundamental nature of learning. Part II treats the topics of maturation, varied activity, repetition, motivation, conditioning, and relationship and organization as six fundamental principles of learning. Part III deals with learning as the improvement of performance and with fatigue and physical handicaps as factors detrimental to learning. Part IV deals with seven forms of learning treated under the heads of the development of motor skills, the development of perception, memorizing, the development of comprehension, thinking as problem-solving activity, the development of emotional activity, and the development of attitudes and ideals. Part V, the concluding section, treats of the retention and transfer of learning.

The book as a whole hews closer to the traditional pattern of texts in educational psychology than certain of the more recent books in this field. In many ways it just misses being an exceptionally well-planned text of the traditional sort. Thus it may be expected to be accorded a mingled reception. Consideration of the psychological aspects of educational development as it relates to the social development of the child and adolescent in our culture tends to be slighted. Many instructors in educational psychology, however, will be just as well content. Such matters, it will be maintained, should be left to texts in genetic and child psychology. But since state education department are notoriously loathe to give certification credit for courses lacking the terms education or educational in their titles, the prospective teacher is all too often denied opportunity of instruction in this important aspect of individual development.

Well-balanced chapters on the development of emotional activity and of attitudes and ideals do much to overcome this deficiency, but in the reviewer's opinion they are not completely adequate. Nevertheless, the book is a scholarly, readable, and current presentation of the sub-

ject which should prove to be a desirable text. The serious student will welcome the comprehensive bibliographies which accompany each chapter.

JAMES M. PORTER, Jr.

Carnegie Institute of Technology.

TRAXLER, ARTHUR E., & TOWNSEND, AGATHA. *Another five years of research in reading*. Educational Records Bulletin No. 46. New York: Educational Records Bureau, 1946. Pp. 192.

Of all the research areas represented by the subjects of the school curriculum, that of reading has been, and still is, the most intensively cultivated. During the five-year period 1940-1945 — and these were “war years” — an average of more than a hundred printed reports of investigations relating to reading appeared annually. In view of this prolific output of research it is important now and then to summarize and to take stock. This the present monograph endeavors to do for the period just noted. Taken together with *Ten Years of Research in Reading* published under the same auspices in 1941, the monograph enables one to note trends with respect to problems studied and techniques employed.

Some of the major trends are recorded in a four-page “Introduction.” In the following sixty-nine pages the research reviewed is classified and discussed under twenty topics. Pages 75-184 contain bibliographical data for 527 “selected” studies together with annotations limited to about one hundred words each. The monograph is concluded by an alphabetical list of the authors of the reading reports and by a subject index.

The contents of the monograph are perhaps best disclosed by the list of topics or problems around which research centered in the period 1940-1945:

1. Reading readiness and beginning reading
2. Reading interests
3. Reading in connection with the school subjects
4. The vocabulary and content of elementary-school readers
5. Vocabulary lists and vocabulary building
6. Phonics
7. Reading tests and testing procedures
8. Speed of reading
9. Eye-movements and reading ability
10. Visual, auditory, and speech defects
11. Dominance, handedness, eyedness and reversals
12. Relationship between reading and other factors
13. Activity programs and reading achievement
14. Hygiene of reading
15. Diagnosis of reading difficulty
16. Remedial and corrective teaching of reading
17. Developmental reading

18. Various aspects of reading (e.g., adult reading)
19. Reading bibliographies and summaries
20. Books and monographs on reading and the teaching of reading.

During the past fifteen years the authors note increased research interest in the problems numbered above as (1), (2), (3), (4), (5), (6), (8), (14), and (17), with some loss of interest in problems (11) and (15) especially.

The authors of the monograph are entirely competent as students of reading, and their comments and criticisms are insightful and suggestive. Consequently, their summary is obviously of signal importance to educators, whether they be concerned with the practical business of teaching children to read in the classroom or with the production of original research. But the monograph should be of value to psychologists as well, for many of the problems discussed are fully as significant psychologically as they are educationally.

WILLIAM A. BROWNELL.

Duke University.

MONCRIEFF, R. W. *The chemical senses*. New York: John Wiley, 1946. Pp. vii+424.

First published in Great Britain in 1944, this book, written by an associate of the Royal Institute of Chemists, covers a wide range of subject matter on taste and smell, taken from the fields of biology, chemistry, electrophysiology, physics, and psychology. Although the author has done an admirable job of coordinating the technical languages of these sciences into a very readable book, one might well ask just what audience or audiences the book was written for. There is no preface disclosing the author's intent.

Of the fourteen chapters the first is devoted to the place of the chemical senses in the animal and human world. Here the author gives a review of the importance of the chemical senses in organisms, ranging from a description of the avoidance of poisons by lower animals to the use of odors in medical diagnosis. The material of four chapters dealing with taste and smell receptors, the nervous system, olfaction, and gustation is very similar to that found in a general psychology text. Chapter 6, which is entitled "The Common Chemical Sense" might well confuse the reader who is unsophisticated in the traditional problems of sense classification. In describing this "separate sense" (p. 130) the author states: "It is clear that the chemical sense is distinct from smell, taste, touch, and pain, and has a separate set of receptors. These receptors are the free nerve endings of spinal and cranial nerves, and are comparable in distinctiveness with heat, cold, or pain receptors, but are more localized." There are various ways of classifying the senses and one of them is based on the stimuli adequate for them. However, this mode of classification has never had the appeal of others. We could speak of the "electrical sense" or the "mechanical sense" or even

of the "sense of noise," but we do not and for good reasons. It would seem better to regard the pain aroused by chemicals as belonging to the same "sense" as that elicited by mechanical, thermal, and electrical stimuli. This chapter describes the common chemical irritants which arouse various sensations, from lachrymators to vesicants. One chapter is devoted to chemical sensibility in lower animals.

The author devotes a chapter to the classification of odors, pointing out the values and limitations of such systems as have been proposed by Rimmel, Zwaardemaker, and Henning. Three chapters on chemical constitution and the physical properties of odorous and taste materials provide an excellent account of the descriptions of chemicals as related to the various sensory experiences they can arouse. In summarizing these relationships into one hundred and twenty "general principles" (e.g., the main factor in determining odour is the architectural type of the molecule) the author states that "they are open to criticism, but the subject is in an early stage of development, and if they are criticized and replaced by a better lot they will have served their purpose." Chapter 12 includes a critical analysis of the theories of odor, with emphasis upon the constitution of odorous particles and the way in which they come into contact with the olfactory apparatus. The salient features of the theories of twenty-three authors are contained in a summary table for purposes of comparison. The author states that "... the transmission of the stimulus from the receptor to the brain is established fact" and adds that the nature of the brain processes is still a mystery. Little is said about theories on the perceptual aspects of odor.

Perfumes and essences, flavor and food, are discussed in the remaining two chapters. Methods are described for separating essential oils for use in perfumes and essences. The chief components of some seventy oils are given in tabular form, showing their various odorous qualities. A detailed chemical analysis of some of the more common oils is given, together with a description of their commercial use. Flavor is regarded as a complex sensation, comprising taste, odor, and touch. Throughout, the author makes effective use of homely illustrations which add "flavor" to the text, such as adding glucose to wines, serving brandy from thin convex glasses, and masking castor oil with peppermint.

For a book which covers such a wide range of subject matter, written primarily around the chemistry of taste and smell as its core, *The Chemical Senses* is well integrated. It might well be placed on a reading list for some advanced courses in psychology. The book is also of value for psychologists as a reference source showing the relationships between chemical stimuli and chemoreception. References to original papers are given in the text, but no summary bibliography is included. The book contains a few charts and tables, numerous chemical formulae, a glossary of technical terms, and a good index.

B. VON HALLER GILMER.

University of Virginia.

Psychological Bulletin

SPONTANEOUS ACTIVITY OF ANIMALS

A REVIEW OF THE LITERATURE SINCE 1929

J. DAVID REED

The Johns Hopkins University

Since the beginning of the century there has been a volume of work studying spontaneous activity of animals, especially the rat and monkey. The best review article is that of Shirley (97) in the *Psychological Bulletin* in 1929. Since that time the subject has not been reviewed comprehensively. Somewhat limited reviews appear in Munn (67), Morgan (66), and Gray (31). Other and even more circumscribed reviews include Richter (73, 74) reporting work done under his direction; Hoskins (40), describing some of the relationships between endocrines and activity; and Kreezer's (52) summary of methods for measuring activity in the rat. A review of diurnal rhythms by Welsh (119) in 1938 discusses much material not directly related to activity. Mettler (64) has reviewed and summarized studies on the effects of striatal injury in 1942.

This article does not attempt to cover work done prior to Shirley's review in THIS JOURNAL. Several of the most important references to work done prior to 1929 are included, but the emphasis has been almost entirely on later material.

THE CONCEPT AND MEASUREMENT OF SPONTANEOUS ACTIVITY

There are two points the reader should keep in mind as he proceeds through the paper. The first is a methodological issue. Much of the research to be reviewed depends on a general concept of spontaneous activity without regard to how the activity is measured. It will become evident in the course of the review that our concept of activity must be tied to the measure of it which we have used, for the results one gets with one measure of activity may be entirely reversed when a different measure is used. The second closely related point is a matter of terminology. Since the largest amount of work has used animals running inside a drum, it will simplify things if the term *activity*, without any qualification, always refers to running activity, not to other measures

Vol. 44 No. 4 July 1947

of activity. Measures other than those in an activity-drum will always be clearly distinguished.

Method and Apparatus

Running Drums. Animals and human beings indulge in spontaneous activity. This observation has been quantified in many ways. The animal most frequently used in experimentation is the rat, whose activity is usually measured in what has been called an activity cage, but will henceforth be referred to as a drum or running drum. This device was first used by Stewart (110) and has been most adequately described by Slonaker (101, 102). It usually consists of two 10-13 inch circular boards mounted on a shaft and separated by a sheet of mesh wound around their periphery (86, 94). The rat runs inside the freely rotating drum, and a counter is attached to record the number of revolutions. Unfortunately, the usual system of measurement has shown only total activity, not activity as a function of time.

Recognizing this inadequacy, Skinner (100) used a Harvard work adder in conjunction with a kymograph to get a summative record whose slope is a constant measure of activity.

The drum has almost as many variations as there have been experimenters in activity. Stewart's 20-inch diameter drum and the 26-inch diameter drum used by Park and Woods (71) represent one extreme, while Shirley (94) used a 10-inch diameter. Results reported in terms of number of revolutions are obviously not comparable when the diameter of the drums is not the same. Furthermore, equating the running by expressing it in distance traversed is of questionable validity in view of Farris' statement that rats in larger wheels run farther than those in smaller wheels (24).

Depending upon the experiment, the rat may live entirely within the drum (94, 95, 101), have a separate living cage, or use supplemental diffuse activity cages (71). Since Richter (73) has shown that the number of revolutions of the drum is reduced when the rat has a choice of several things to do, the results of different experimenters may not be comparable.

The revolving drum has been the most extensively used laboratory instrument in investigating activity. Its physical variables have been discussed by Skinner (100) and Lacey (53). The reliability of the measures obtained is remarkable—Shirley (94) reports a rank-order correlations of .97 for five-day totals of activity, and a split-half r of .90. Beach's (4) figures are even higher (.98). Unfortunately a basic assumption in these results involves the equivalence of the measures. Lacey

(53) raises the very justifiable criticism that the measure may be showing only the consistency of the different drums. It is significant to note that in one case in which the animals were changed from one cage to another, the correlation reported was .80 (113). There are wide individual differences even between litter mates in normal rats with respect to running, some rats running 200 revolutions per day and others 20,000. The pattern of running is set up by the tenth day or not at all. After this time the individual differences are relatively constant.

The running drum has been used to indicate tension or motivation in the rat. Thus, Durrant (20) and Slonaker (106) have correlated running with sex drives. Geier and Tolman (28, 29) have used running behavior to indicate increase in tension in the rat.

Dorcus (16) devised a cage which moved slowly toward a goal object when the rat ran inside of it.

Tambour- or Spring-Mounted Cages. Another apparatus for measuring activity is that first used by Syzmanski (114, 115) which consisted of a spring-mounted cage attached to a lever recording system. The disadvantage of lack of damping has been somewhat overcome by tambour-mounted activity cages (73, 109). The three supporting tambours are joined to one tube and record every movement on a kymograph. Both these methods produce records according to time, but records which are difficult to treat quantitatively because no ready means of determining the total activity is available.

Hunt and Schlosberg (42, 43) counted the number of 5-minute active periods occurring in such a cage over varying intervals of time, and Irwin (44) recorded the number of active seconds per minute in newborn children. Wilbur (121) used a spring mounted cage connected to a Harvard work adder to obtain a summative record (of the activity of chicks) which is much easier to interpret.

Smith (109), measuring audiogenic and electrogenic convulsive activity, supported a cage from four pneumographs or by one large flexible hydron bellows. Oscillation could be reduced by means of a small vent.

Other animals have been used with appropriately modified cages to record activity. Monkeys have been fastened by a nine-inch chain to a 2.5-inch rod, so that movement caused the rod to advance a counter (84). A monkey-sized pneumatically-mounted activity cage has been used by Kennard, Massimy and Chevallier (51, 62).

Other Automatic Methods of Recording Activity. Another measure of activity has been suggested, incorporating a tilting box (7) in which the movement of the rat from one end of the box to the other advanced a counter. Claiming that the tilting motion would interfere with accurate

measurement of rat's activity, Siegel (98) utilized the animal's motion from one end of a 22×6 inch box to the other to break a photoelectric relay and thus advance a counter.

A horizontal turntable for exercising rats has also been used to record activity (21, 22). Since the distance the rat runs depends upon its proximity to the center of the turntable, it is probable that this device will not be popular in controlled experiments.

Curtis (15), working under Liddell, reports the use of a pedometer to record activity of the sheep and the pig. Head-shaking in chickens has also been reported by Levy (58).

Observational Method of Quantifying Activity. An observational means of recording activity has been used by Hall (32), Beach (3), and Fredericson (25). Hall recorded the distance traversed by rats in a round open field eight feet in diameter. Beach noted which of 36 squares a rat entered upon in the ten minutes it was free in an area three feet square. Fredericson observed six classes of behavior indulged in by rats in a field two feet square.

Spontaneous Activity as a Behavior Category

Now let us take a moment to see what the methods just reviewed have to do with the concept of activity. Most of the authors tend to lump all manifestations of activity together and to pin one label on all of them—*activity*. This failure to distinguish types of activity in terms of its measure leads to a false concept of activity, for what data we have point to more than one type, or at least more than one aspect of activity.

There are, for example, wide individual differences in the running activity of rats but not nearly such wide differences in restless cage activity. Tainter (116) found that caffeine, metrazol, and picrotoxin had no effect on running but did increase behavior measured in a diffuse activity cage. Hunt and Schlosberg (43) found only 9% decrease in diffuse activity with castration instead of the 98% found by Hoskins (40) for running activity.

In light of these considerable differences, it seems logical that different terms should be used to distinguish the two devices and the behaviors which they measure. Throughout this paper, the author has attempted to distinguish between running activity in the rotating drum, on the one hand, and diffuse activity in a cage or stabilimeter on the other. As long as the terminology for these two distinct situations is the same, the notion will tend to persist that they are strictly comparable measures, which they are not.

HEREDITY AND AGE

Genetic Basis. Rundquist (89) by selective breeding has been able to get active and inactive strains of rats. The active strain is less easy to purify than the inactive strain. Selection for running produced strains in which there were measurable: (1) increases in number of successful matings, (2) increases in sizes of litters, and (3) decreases in the gestation period. The active rats also had a higher basal metabolic rate than the inactive strain. The selective breeding of these strains has been carried through 29 generations, with no change beyond the 12th (9, 30). Brody has concluded that the two strains differ with respect to a single gene which acts as a dominant in males and a recessive in females. This gene must act as an inhibitor, since none of the matings within the inactive strain produces active offspring, but on the other hand, active-strain matings produce individuals which vary from extreme inactivity to extreme activity. The genetic factors are somewhat obscured by environmental influences.

Age. Running activity of rats increases with age until the animals are about 80 days old, then is relatively constant until about 120 days, after which it gradually falls off till death (72, 95, 101). Richter (73) determined the amount of running, diffuse activity, and nest building as a function of age. The more active rats tend to have shorter lives than the inactive rats.

THE INTERNAL ENVIRONMENT

Nutrition

The daily running activity of the rat increases just prior to the normal time of feeding, even though the rat has been fed in what would otherwise be an inactive period (72, 103). This fact is probably explainable on the basis of hunger or metabolic changes connected with the daily 24-hour hunger rhythm initiated by constant feeding at a specific hour (120). Studies made by Richter (73) on generalized activity, however, show that the rat very probably has a two-hour hunger rhythm if he is allowed to have food constantly available.

If a rat is deprived of food for a period of time, its activity will tend to increase for as much as 96 hours. If deprived of food and water, it increases for 72 hours before it drops off (73, 117), probably due to weakness. Brobeck (8) showed that if food intake and environmental temperature are held constant, there is a negative correlation between activity and weight gain. Smith and Conger (107) varied the diet of rats by keeping the caloric value constant but changing the proportion

of fat or protein. They found that up to 56% of the caloric value of food may come from fat without reduction in spontaneous activity. Fifty per cent of animal protein, however, induces marked reduction in running. Following ingestion of protein there is a marked metabolic rise which is not present following ingestion of fats. This result extends Slonaker's (105) finding that a diet of 14 to 18% protein yields maximal spontaneous running. Hitchcock (39), however, fed rats eight grams of meat daily and observed little effect on running.

A protein-free diet for short periods of time is effective in increasing activity. But a low-protein diet over a long period of time depressed activity (39).

When allowed to select the amount of alcohol ingested, rats maintained their activity, although the forcing of alcohol by inclusion in the drinking water supply caused a reduction in activity (35).

Vitamins

Rats deprived of food, water, food and water, thiamin, and riboflavin showed increased activity until they were given ample amounts of the formerly scarce substance; then they were quiescent to a greater degree than normal until they had again 'caught up' (117). Limitation of Vitamin B is promptly followed by increased running activity in the rat; after five to ten days, however, activity is sharply decreased, long before any clinical signs of deficiency appear. This is accompanied by a drop in quantity of food ingested. When these rats are then given unlimited amounts of food and vitamins, activity decreases and remains low throughout the period of increased intake (5). These two papers do not completely corroborate Jackway's earlier finding (47) that slight deprivation of Vitamin B complex, not severe enough to cause any symptoms except retardation of growth and slight roughness of coat, results in a decrease in voluntary activity.

In a very suggestive paper Ziegler and Knudsen (124) showed that young white rats fed on a rachitic diet during infancy ran less after recovery from the rickets than normals. If, however, the rats were deprived of Vitamin D at an even earlier age by feeding the rachitic diet to their mothers during pregnancy, the surviving pups were more active after recovery than normals.

That deficiency is not always accompanied by increased activity is demonstrated by the fact that rats deprived of the magnesium ion show a consistent drop in activity without the initial rise usually accompanying deprivation (117). Smith and Smith (108) observed that rats fed a diet low in inorganic constituents gradually declined in vol-

untary running. The reduction in running occurred long before the appearance of impaired running ability.

Drugs

There has been interest in the effects of drugs on behavior as showing perhaps quantifiable action similar to that qualitatively produced in human beings. In general, analeptics stimulate activity (6), but have a depressant or diminishing effect with prolonged use. Thus cocaine, benzedrine, ephedrine, and propadrine increased running for single doses, but effects of tolerance showed when the specific drug was administered for several days (116, 123). Caffeine, metrazol, and picrotoxin depressed running activity, but enhanced restless activity as measured in diffuse activity cages (116). In a subsequent paper, Shulte and Tainter (91) have shown the differential temporal effects of the administration of caffeine, coramine, and metrazol on diffuse activity.

Running activity is increased by benzedrine sulphate and by Kola, but there is a decrement in activity just before feeding time 24 hours later (35, 92).

Phenobarbital administered intraperitoneally (69) to white rats depressed the nighttime running level, but had no effect on the daytime level which is normally much lower than that of the nighttime. These injected rats ate less, drank less, and gained less weight. Even after withdrawal of phenobarbital, the experimental rats were less active than the controls (49).

Other drugs used have been dinitrophenol and thiouracil. These are discussed below.

Endocrines

Adrenals. In a review of experiments dealing with endocrines up to 1927, Hoskins (40) reports that entire removal of the adrenals leads to marked reduction in running. This operation is very difficult to perform successfully, but in all cases in which there was little reduction in activity, subsequent autopsy showed hypertrophied adrenal particles. In 12 out of 27 adrenalectomized rats, Richter (79) was able to restore running partially by implants of adrenal gland to the ovaries. Subsequent removal of the grafts decreased activity again. Histological examination showed viable cortical tissue but no medulla. Increase of salt in the diet produced a definite but partial increase in activity. Ergographic studies show that while the absolute strength of muscles remains unaltered, the capacity for work following adrenalectomy is reduced to 8% of normal. Bilateral abdominal sympathectomy and

adrenal inactivation by section of the nervous connection reduced running activity slightly for a period of 10 days with no apparent effect after that time (1). Lacey (54) injected adrenalin into rats and found a decrease in total diffuse activity although restlessness as indicated by behavior in the normally inactive period increased markedly.

Thyroid. Hoskins (40) has reported no reduction in running resulting from thyroidectomy. This result has been questioned by Hall and Lindsay (34) who found 50% reduction using the same measure as that reported by Hoskins. Richter (75) has pointed out that the results obtained depend upon the amount of thyroid tissue remaining, for a very small piece is enough to sustain activity. Thyroidectomized rats may be re-activated by replacement therapy. However, if the quantity of thyroid administered is too great, the activity is once more reduced. That reduction in metabolic rate as such is not responsible for decrease in activity was shown by increasing the metabolism of thyroidectomized and normal rats by dinitrophenol (33). Inhibition of the thyroid by thiouracil resulted in slower growth, lowered metabolism, increase in the length of interoestral periods, reduction of spontaneous activity and disruption of rhythmic patterns of activity (61). Rundquist's breed of active rats had a higher metabolic rate than his inactive strain (90).

Pituitary. Ablation of the pituitary leads to marked decrease in running activity. Injection of emulsion of fresh anterior lobe of the hypophysis reduced the decrease in activity, but only in one animal out of seven did this replacement therapy cause activity to remain at the original level. Viable hypophyseal transplants to the anterior chamber of the eye had only a slight stimulating effect (81).

When the stalk of the pituitary is severed in the female rat, the normal oestral periodicity is greatly slowed down. Five female rats showed oestral periods separated by the following numbers of days: 9.6, 11.2, 13.0, 14.3, 18.4. Richter explains this as being probably a whole multiple of the fundamental four-to-five-day ovulation cycle. This explanation is corroborated by the fact that ovariectomy eliminated all cycles (77).

The genetically hypopituitary animal, the dwarf rat, possesses a diurnal rhythm (70).

Liver. Normal function of the liver is important in maintaining activity. Ligation of the bile duct caused almost complete reduction in animals in which the duct remained occluded. Associated with this effect was widespread destruction of liver cells, considerable dilation of the bile duct, and other profound peritoneal effects (80). Partial hepatectomy was performed by Dugal and Ross (17) who found that the

25% of the liver remaining took the same length of time to regenerate as did activity to return to normal.

Pancreas. The decrease in activity following pancreatectomy is variable. Some rats show little loss, others great. This reduction is attributed to the inability to metabolize carbohydrates (85).

Ovaries. There are several measurable sex differences in daily running activity. The female goes through a cycle of activity of about four to five days' duration, the peak of which is reached at oestrus (23, 118). Hourly records of the female show variation in the distribution of the activity as a function of the time in the oestrus cycle (103). This cycle may be abolished by bilateral ovariectomy (77, 87), when the activity in the drum is decreased by as much as 98%, but it is restored by ovarian transplants (87). The presence of the uterus is not essential to this activity cycle, as is shown by the fact that hysterectomy has no permanent effect on running activity (18, 19). If copulation takes place during oestrus, the activity is reduced until pregnancy and lactation are over and the female is again in oestrus as shown by cornified epithelial cells or receptivity. Stimulation of the cervix during receptivity leads to pseudo-pregnancy with about a 10-day marked reduction in running, followed by gradually increasing activity until the previous level is reached in about six days (104).

The cycles of the ovariectomized rat may be restored in females or impressed in castrated males by several means, all of which involve supplanting the hormonal deficiency. Pregnancy urine of women seven to nine months pregnant, given in drinking water, was found to restore activity to near normal levels (78). Amniotin injected in spayed females and castrated males had a similar effect (82). The earlier discrepancy between results of other investigators, reported by Shirley, is possibly explicable by the findings of Young and Fish (122). They were able to restore activity to levels prevailing before gonadectomy by estrone administered in such a way that a constant source of the hormone was maintained. A certain small amount of estrone is necessary for the manifestation of running activity, but beyond that quantity, further estrone does not increase the activity nor alter its periodicity. These are similar to Heller's results (36) on male hormone. Castrated males show oestrus cyclic activity from ovarian transplants. Elderly males increase activity and lose weight following ingestion of an estrogenic fraction of chorionic gonadotropic extract (41).

Testes. The running activity of the male is generally much less than that of the female, and lacks its cyclical characteristics. Complete castration leads to marked reduction (about 98%) in running. Fractional

castration by Gans (27) of the following amounts of the testes of rats: 0, 1, $1\frac{1}{2}$, $1\frac{3}{4}$, 2, led to the following decrements in running compared with his controls: 20%, 11%, 26%, 48%, 79%. Probably there was not time for regeneration to take place. Castration before the 12th day after birth leads to no activity reduction in the adult rat (26) but this statement has been questioned by Richter (76) who found the reduction in running almost as great as in rats castrated after sexual maturity. Richter ascribes this difference to the relative inactivity of Gans' controls. It is possible, however, that the findings of Hunt and Schlosberg (42, 43) would bear on the question. They used general diffuse activity cages and found little difference between the number of five-minute active periods of normal and castrated males. The running activity of both male and female castrated rats is increased by testicular grafts, provided these grafts are "functional takes" (87).

The relation between sexual drive and running activity is unclear. A male rat in a running drum near one female rat will show peaks of activity correlated with her oestral periods. If several females are in nearby drums, the total running of the male is increased but is no longer of the simple oestral periodicity (20, 106). On the other hand, the correlation between sex drive (as measured by copulatory tests and the Columbia obstruction box) and running activity is near zero (113). Rundquist (89) found that rats bred for activity were more fertile than those bred for inactivity. These facts are not contradictory but do require clarification.

EXTERNAL ENVIRONMENT

Light and Darkness

The rat displays more running and more diffuse activity in darkness than in light. This rhythm is of a 24-hour periodicity and is remarkably stable. The daily activity periods are determined by the internal rhythm of the animal, which rhythm is only gradually changed by the external conditions—light and temperature (10, 13, 14, 43, 48). Using a rhythm involving alternation of eight hours of dark and eight hours of light, Hemmingsen and Krarup were unable to abolish the 24-hour rhythm even in rats kept under these conditions from birth (37). By using six hours of light and warmth, alternating with six hours of cooler darkness, Browman (14) was able to impress a 12-hour rhythm on about 30 out of 32 rats. The rhythm was not stable, however, as shown by the fact that the longest period during which it persisted in any one rat was 37 days, in spite of the continual 12-hour cycle of light and heat.

The female rat shows a four-to-five-day oestrous cycle of activity

which can be somewhat altered by lighting in that constant light leads to longer intervals between peaks of oestrous activity and lowered daily activity (37). Browman (10) even found evidence of constant cornification (but not receptivity) in some of his animals. Subsequently he demonstrated that the intact visual apparatus was necessary for this shift to take place (11, 12). Females who were blinded showed the same 24-hour rhythm as those reared in constant darkness. Feeding schedules apparently had little effect on the running time. That the intensity of the constant light is important has been nicely demonstrated by Johnson (48) who found that the amount of change in the period of maximal activity of mice was related to the quantity of cage illumination. These results were obtained with mice in diffuse activity cages, since the mouse apparently exhibits even more dichotomous activity cycles than the rat. Working with male rats in diffuse activity cages, Hunt and Schlosberg (43) attempted to determine the source of the 24-hour activity rhythm; since castration did not abolish the periodicity they came to the conclusion that it was not based on the testes. Reversing the normal day and night periods caused the rats to shift their period of maximal activity in about a week.

In constant light, the female rat will show peaks of activity which are correlated with a daily cool period (13). This is also true of blinded rats. The pre-oestrus activity readings come at the end of the 12-hour cool period. It had previously been shown that in a hot room, rats will reduce their activity slightly (13, 97). Below about 40 degrees Fahrenheit the rat's activity is reduced also.

Temperature

The effect of temperature on the diffuse activity of two-day-old mice was determined by Stier (111). These young mice are still poikilothermic so that control of external temperature had direct effect on the body temperature. It was found that different straight-line Arrhenius plots would fit the several measures used: amount of activity, quiescence, and frequency of occurrence. Thus, he concluded that activity is a function of several factors.

Activity Following Confinement

The drive for activity hypothesized in several behavior theories has not been fully substantiated. Shirley (95) reports increased running activity after a rat has been confined one to two days, but after ten days of enforced activity, running decreases. Siegel (99) found an increase in the number of interruptions of a light beam following close confinement.

It is important to note that in neither case is the amount of activity under conditions of no confinement significantly different from the increase reported after confinement.

Skinner (100) has stated that "if any extensive activity is prohibited during part of a day, the remaining part shows a greater 'density' of activity per unit of time." Since he measured activity for a five to six hour period beginning at 3:00 a.m., it is possible that his conclusion is an artifact based upon the normal peak of the rat's running activity which occurs at about 2:00 a.m. After this time there is a gradual decrease in the amount run, a rhythm which will persist in constant darkness. Without further work substantiating the generality of the principle of greater activity following inactivity, it would be parsimonious to ascribe Skinner's results to the specific hours at which he recorded activity.

Miscellaneous Studies

Other studies using the running drum which might be mentioned are: the correlation between activity and errors in learning a maze is low (89, 96); while that between activity and time to traverse mazes is higher (68); rats given difficult discriminations run less than rats that are not made abnormal by difficult problems (60); a series of electroconvulsive shocks greatly reduced voluntary activity in rats (112).

NEURAL CONTROL OF SPONTANEOUS ACTIVITY

The search for a neural center which controls activity has led to contradictory evidence. There is complete agreement that frontal lesions do increase activity and that bilateral lesions are more effective than uni-lateral lesions. The hyperactivity usually takes two to three weeks to emerge. The animals on which most of the work has been done are rats, cats, and monkeys.

The first observation of the effect of cortical lesions on activity has been attributed to several clinicians and experimenters. In 1920 Lashley (57) quantified the change in activity of the rat by using a running drum and noted that only frontoparietal lesions both increased the number of hours of running and decreased the time spent in resting. Jacobsen (45, 46) noted increased restlessness and general activity following frontal destruction in the monkey, and Langworthy and Kolb (56) described the behavior of cats with heightened restlessness. Since 1937, considerable evidence has been gathered concerning the effect on spontaneous activity produced by lesions in the brain.

Rats

In the rat, the unilateral removal of the frontal pole did not appear to augment running activity. The inactive rats became hyperactive while the active rats did not increase their running relatively so much. Bilateral ablation of the frontal poles was much more effective in increasing running than the unilateral (3, 83). Beach (4) measured running for 30 days before and 50 days after electrolytic destruction of varying amounts of the corpus striatum. Activity increased postoperatively in one animal, decreased in two, and was unchanged in two others. "No relationship between the magnitude of lesion and effects on activity could be determined. In the rat, the striatum evidently does not exert a controlling effect upon running activity as measured in this experiment." On the other hand, Richter and Hines (84) found that monkeys with unilateral striatal lesions immediately had greatly increased activity, and Mettler (64) reports hyperactivity in cats following striatal lesions.

According to Mettler (63, 64) when the striatum is injured, hyperkinesia is the rule. He asserts that the striatum is an inhibitory mechanism; "... stimulation of it produces inhibition and removal of it engenders evidence of motor release. It stands on the one hand between the cortex and the final common path as part of the route through which the cortex may exert an inhibitory effect and, on the other hand, it operates between the thalamus and lower motor mechanisms in the automatic inhibition incident to 'unconscious activity'." If the cerebral cortex is totally removed, the decorticated animal does not exhibit incessant activity but shows an inability to initiate or inhibit movement suddenly (65).

Cats

Cats with bilateral one-stage removal of the rostral portions of the cerebral hemispheres were noted by Magoun and Ranson (59) to be almost continually walking about. Langworthy and Richter (55) recorded increase in activity from 27 to 61 units from unilateral operation and to 399 units for the bilateral removal of motor cortex, premotor cortex, and possibly a small tip of the corpus striatum in cats.

Monkeys

In monkeys, Richter and Hines (84) found that bilateral removal of areas 8, 10, 11, and 12 had little effect on activity, while that of 9 did (62). On the other hand, Kennard and Ectors (50) reported increased activity following removal of area 8 alone. These results are not con-

tradictory in view of the method of measurement of activity. Richter and Hines (84) attached the monkey by a chain to a short steel rod projecting from an axle. Movement of the monkey caused the rod to advance a counter. The method of recording activity used by Kennard *et al.* was a pneumatically mounted diffuse activity cage.

Generally the activity builds up in the course of two to three weeks following an operation. Ruch and Shenkin (88), however, report that lesions in area 13 (of Walker) consistently produce hyperactivity within the second post-operative day. Richter and Hines (84) also report such immediate hyperkinesia when the monkey has striatal lesions.

Kennard *et al.* (51) have stressed the visual role in hyperactivity in monkeys. "Hyperactivity is markedly affected by visual stimuli. It disappears in the dark or when the animals have been deprived of vision either by enucleation of the eyes or by bilateral lobectomy. Absence of auditory stimuli has not the same effect."

A decrease in activity was noted by Barris (2) following "bilateral one-stage removal of the rostral portions of the neo-cortex of cats." Kennard *et al.* reported that "hypermotility in monkeys and chimpanzees is related to lesions of the rostral portions of area 6 and to area 8" (51).

SUMMARY

The literature of the last twenty years concerning activity in animals has been reviewed. The methods, results and concepts of activity have been summarized and appraised.

Several methods have been in use:

1. The running drum: this apparatus yields high reliability for measures taken on a particular drum, but there are also inconsistencies from one drum to another and from one experimenter's design of drum to another.
2. The diffuse activity cage: a cage mounted on tambours or springs. The record which it gives varies widely from one cage to another.
3. Several miscellaneous mechanical and observational methods, which have not been extensively used.

There are considerable individual differences in running activity as well as variability of one animal's activity from time to time. The individual differences are due in part to heredity, but are somewhat complicated by environmental influences. Intra-animal variability can be ascribed to several factors: running increases during hunger and most deprivations, during darkness and cool periods, and during oestrus. In various kinds of endocrine imbalance or deficiency, there is usually a decrease in activity—a marked decrease in the running drum but only a small decrement in the diffuse activity cage.

Running activity and diffuse activity are sometimes affected in the same way, sometimes differentially. Both reach a maximum during the cool or dark part of a 24-hour cycle. Some of the analeptic drugs stimulate both kinds of activity, but other drugs may increase diffuse activity while decreasing running activity.

Injury to the brain affects activity. In particular, lesions of the frontal cortex heighten activity, and bilateral lesions cause a greater increase than unilateral injury. Still, it is not yet clear whether there is a specific activity center in the brain and, if so, where it is.

There is now a very large body of data concerning animal activity, but it needs further definition and interpretation. Particularly needed is a clarification of the concept of activity in relation to the method of measuring it. Most treatments of the subject tend to regard activity as a single entity. Yet, in some cases, where comparable measures of activity are available from different devices, running drum and diffuse activity cage, the results are not the same. Activity, it would then appear, does not constitute a single behavior category which can be measured with any instrument but must be considered, for the present at least, in terms of its method of measurement.

BIBLIOGRAPHY

1. BACQ, Z. M. The effects of abdominal sympathectomy, adrenal inactivation and removal of the stellate ganglia on the spontaneous activity of the albino rat. *Endocrinology*, 1931, **15**, 34-40.
2. BARRIS, R. W. Cataleptic symptoms following bilateral cortical lesions in cats. *Amer. J. Physiol.*, 1937, **119**, 213-220.
3. BEACH, F. A. Effects of brain lesions upon running activity in the male rat. *J. comp. Psychol.*, 1941, **31**, 145-179.
4. BEACH, F. A. Effects of lesions to corpus striatum upon spontaneous activity in the male rat. *J. Neurophysiol.*, 1941, **4**, 191-195.
5. BLOOMFIELD, A., & TAINTER, M. L. The effect of vitamin B deprivation on spontaneous activity of the rat. *J. Lab. clin. Med.*, 1943, **28**, 1680-1690.
6. BOUGHTON, L. L. The effect of life cycle therapeutic dosage administration of drugs to albino rats. II. On activity, maze learning and re-learning. *J. Amer. pharm. Ass.*, 1942, **31**, 240-244.
7. BOUSFIELD, W. A., & MOTE, F. A. The construction of a tilting activity cage. *J. exp. Psychol.*, 1943, **32**, 450-451.
8. BROBECK, J. R. Effects of variations in activity, food intake and environmental temperature on weight gain in the albino rat. *Amer. J. Physiol.*, 1945, **143**, 1-5.
9. BRODY, E. G. The genetic basis of spontaneous activity in the albino rat. *Comp. Psychol. Monogr.*, 1942, **17**, No. 5. Pp. 24.
10. BROWMAN, L. G. Light in its relation to activity and estrous rhythms in the albino rat. *J. exp. Zool.*, 1937, **75**, 375-388.
11. BROWMAN, L. G. The effect of bilateral optic enucleation on the voluntary muscular activity of the albino rat. *J. exp. Zool.*, 1942, **91**, 331-344.
12. BROWMAN, L. G. The effect of bilat-

- eral optic enucleation upon the activity rhythms of the albino rat. *J. comp. Psychol.*, 1943, 36, 33-46.
13. BROWMAN, L. G. The effect of controlled temperatures upon the spontaneous activity rhythms of the albino rat. *J. exp. Zool.*, 1943, 94, 477-489.
 14. BROWMAN, L. G. Modified spontaneous activity rhythms in rats. *Amer. J. Physiol.*, 1944, 142, 633-637.
 15. CURTIS, Q. F. Diurnal variation in the free activity of sheep and pig. *Proc. Soc. exp. Biol. & Med.*, 1937, 35, 566-567.
 16. DORCUS, R. M. A new device for studying motivation in rats. *J. comp. Psychol.*, 1934, 18, 149-151.
 17. DUGAL, L. P., & ROSS, S. Effet de l'ablation partielle du foie sur l'activité spontanée du rat blanc. *Rev. Canad. Biol.*, 1943, 2, 435-441.
 18. DURRANT, E. P. Studies on vigor. XI. Relation of hysterectomy to voluntary activity in the white rat. *Amer. J. Physiol.*, 1927, 82, 14-18.
 19. DURRANT, E. P. Relation of hysterectomy of long standing to voluntary activity in the white rat. *Amer. J. Physiol.*, 1931, 97, 519. (Abstr.)
 20. DURRANT, E. P. Influence of the female white rat in bodily activity of the male. *Amer. J. Physiol.*, 1935, 113, 37. (Abstr.)
 21. FARRIS, E. J., & ENGUALL, G. Turntable for exercising rats. *Science*, 1939, 90, 144.
 22. FARRIS, E. J. Apparatus for recording cyclical activity in the rat. *Anat. Rec.*, 1941, 81, 357-361.
 23. FARRIS, E. J. Leucopenia associated with normal estrous in the albino rat. *Anat. Rec.*, 1942, 82, 147-151.
 24. FARRIS, E. J. Breeding of the rat. Ch. 1 in J. Q. Griffith, Jr. and E. J. Farris (Eds.), *The rat in laboratory investigation*. Philadelphia: Lipincott, 1942. Pp. 1-17.
 25. FREDERICSON, E. The theory of psychomotion as applied to a study of temperament. *J. comp. Psychol.*, 1946, 39, 77-89.
 26. GANS, H. M. Studies in vigor. XIII. The effect of early castration on the voluntary activity of male albino rats. *Endocrinology*, 1927, 11, 141-144.
 27. GANS, H. M. Studies on vigor. Effect of fractional castration on the voluntary activity of male albino rats. *Endocrinology*, 1927, 11, 145-148.
 28. GEIER, F. M. The measurement of tension in the rat. A contribution to method. *J. comp. Psychol.*, 1942, 34, 43-49.
 29. GEIER, F. M., & TOLMAN, E. C. Goal distance and restless activity. I. The goal gradient of restless activity. *J. comp. Psychol.*, 1943, 35, 197-204.
 30. GRAVES, E. A. The genetic basis of activity in the albino rat. *Psychol. Bull.*, 1937, 34, 757-758. (Abstr.)
 31. GRAY, W. L. *The effects of forced activity on maze learning and the selection and consumption of food by rats*. Unpublished Ph.D. Thesis, Johns Hopkins Univ., 1933.
 32. HALL, C. S. Emotional behavior in the rat. III. The relationship between emotionality and ambulatory activity. *J. comp. Psychol.*, 1936, 22, 345-352.
 33. HALL, V. E., & LINDSAY, M. The effect of dinitrophenol on the spontaneous activity of the rat. *J. Pharm. exp. Therap.*, 1934, 51, 430-434.
 34. HALL, V. E., & LINDSAY, M. The relation of the thyroid gland to the spontaneous activity of the rat. *Endocrinology*, 1938, 22, 66-72.
 35. HAUSMANN, M. F. The behavior of albino rats in choosing food and stimulants. *J. comp. Psychol.*, 1932, 13, 279-309.
 36. HELLER, R. E. Spontaneous activity in male rats in relation to testis

- hormone. *Endocrinology*, 1932, 16, 626-632.
37. HEMMINGSEN, A. M., & KRARUP, N. B. Rhythmic diurnal variations in the oestrus phenomena of the rat and their susceptibility to light and dark. Det. Kgl. Danske Videnskabernes Selskab., *Biologiske Meddelelser*, 1937, 13, 1-61.
38. HERRING, V. V., & BRODY, S. Growth and development. XLIII. Diurnal metabolic and activity rhythms. *Univ. Missouri Agri. Exp. Station Res. Bull.* 274, 1938 (not seen).
39. HITCHCOCK, F. A. The effect of low protein and protein-free diets and starvation on the voluntary activity of the albino rat. *Amer. J. Physiol.*, 1928, 84, 410-416.
40. HOSKINS, R. G. Studies on vigor. XVI. Endocrine factors in vigor. *Endocrinology*, 1927, 11, 97-105.
41. HOSKINS, R. G., & BEVIN, S. The effect of fractionated chorionic gonadotropic extract on spontaneous activity and weight of elderly male rats. *Endocrinology*, 1941, 27, 927-931.
42. HUNT, J. McV., & SCHLOSBERG, H. General activity in the male white rat. *J. comp. Psychol.*, 1939, 28, 23-38.
43. HUNT, J. McV., & SCHLOSBERG, H. The influence of illumination upon general activity in normal, blinded and castrated male white rats. *J. comp. Psychol.*, 1939, 28, 285-298.
44. IRWIN, O. C. Effect of strong light on the body activity of newborns. *J. comp. Psychol.*, 1941, 32, 233-236.
45. JACOBSEN, C. F. A study of cerebral function in learning. The frontal lobes. *J. comp. Neurol.*, 1931, 52, 271-340.
46. JACOBSEN, C. F. Studies of cerebral function in primates. *Comp. Psychol. Monogr.*, 1936, 13, No. 3. Pp. 68.
47. JACKWAY, I. Voluntary activity in the rat as related to the intake of whole yeast. *J. comp. Psychol.*, 1938, 26, 157-162.
48. JOHNSON, M. S. Effect of continuous light on periodic spontaneous activity of white-footed mice. *J. exp. Zool.*, 1939, 82, 315-328.
49. JONES, M. R. The effect of phenobarbital on food and water intake, activity level, and weight gain in the white rat. *J. comp. Psychol.*, 1943, 35, 1-10.
50. KENNARD, MARGARET A., & ECTORS, L. Forced circling in monkey following lesions of the frontal lobes. *J. Neurophysiol.*, 1938, 1, 45-54.
51. KENNARD, MARGARET A., SPENSER, SUSAN, & FOUNTAIN, G., JR. Hyperactivity in monkeys following lesions of the frontal lobes. *J. Neurophysiol.*, 1941, 4, 512-524.
52. KREEZER, G. L. Technics for the investigation of psychological phenomena in the rat. Ch. 10 in J. Q. Griffith, Jr. and E. J. Farris (Eds.), *The rat in laboratory investigation*. Philadelphia: Lippincott, 1942. Pp. 199-273.
53. LACEY, O. L. A revised procedure for the calibration of the activity wheel. *Amer. J. Psychol.*, 1944, 57, 412-420.
54. LACEY, O. L. The dependence of behavior disorders in the rat upon blood composition. *J. comp. Psychol.*, 1945, 38, 277-284.
55. LANGWORTHY, O. R., & RICHTER, C. P. Increases in spontaneous activity aroused by frontal lobe lesions in cats. *Amer. J. Physiol.*, 1939, 126, 158-161.
56. LANGWORTHY, O. R., & KOLB, L. C. The experimental production in the cat of a condition simulating pseudo-bulbar palsy. *Amer. J. Physiol.*, 1935, 111, 571-577.
57. LASHLEY, K. S. Studies of cerebral function in learning. II. *Psychobiology*, 1920, 2, 55-136.
58. LEVY, D. M. On the problems of

- movement restraint, tics, stereotyped movements, hyperactivity. *Amer. J. Orthopsychiat.*, 1944, **14**, 644-671.
59. MAGOUN, H. W., & RANSON, S. W. The behavior of cats following bilateral removal of the rostral portion of the cerebral hemispheres. *J. Neurophysiol.*, 1938, **1**, 39-44.
 60. MAIER, N. R. F., & WAPNER, S. Studies of abnormal behavior in the rat. *J. comp. Psychol.*, 1944, **37**, 151-158.
 61. MANN, C. W. The effect of thiouracil upon the heart rate, estrous cycle and spontaneous activity of the white rat. *J. Psychol.*, 1945, **20**, 91-99.
 62. MASSIMY, R., & CHEVALLIER, R. J. Les effets, chez le singe, de l'ablation préfrontal unilatérale; modifications de l'activité, du mode réactionnel et des réflexes. *C. R. Soc. Biol. Paris*, 1942, **136**, 103-106.
 63. METTLER, F. A. Relation between pyramidal and extra-pyramidal function. *Res. Publ. Ass. nerv. ment. Dis.*, 1942, **21**, 150-227.
 64. METTLER, F. A., & METTLER, C. C. The effects of striatal injury. *Brain*, 1942, **65**, 242-255.
 65. METTLER, F. A., METTLER, C. C., & CULLER, E. A. Effects of total removal of cerebral cortex. *Arch. Neurol. Psychiat.*, Chicago, 1935, **34**, 1238-1249.
 66. MORGAN, C. T. *Physiological psychology*. New York: McGraw-Hill, 1943.
 67. MUNN, N. L. *An introduction to animal psychology: the behavior of the rat*. New York: Houghton Mifflin, 1933. Pp. 50-78.
 68. OMWAKE, L. The activity and learning of white rats. *J. comp. Psychol.*, 1933, **16**, 275-285.
 69. OMWAKE, L. The influence of barbitol on the activity and learning of white rats. *J. comp. Psychol.*, 1933, **16**, 317-325.
 70. OSBORN, C. M. Spontaneous diurnal activity in a genetically hypopituitary animal, the dwarf rat. *Anat. Rec.*, 1940, **78**, Suppl. p. 137.
 71. PARK, O., & WOODS, L. P. A modified Hemmingsen-Krarup mammalian activity recorder. *Proc. Soc. exp. Biol. Med.*, 1940, **43**, 366-370.
 72. RICHTER, C. P. A behavioristic study of the activity of the rat. *Comp. Psychol. Monogr.*, 1922, **1**, 1-55.
 73. RICHTER, C. P. Animal behavior and internal drives. *Quart. Rev. Biol.*, 1927, **2**, 307-343.
 74. RICHTER, C. P. Symposium: Contributions of psychology to the understanding of problems of personality and behavior. IV. Biological foundations of personality differences. *Amer. J. Orthopsychiat.*, 1932, **2**, 345-354.
 75. RICHTER, C. P. The role played by the thyroid gland in the production of gross body activity. *Endocrinology*, 1933, **17**, 73-87.
 76. RICHTER, C. P. The effect of early gonadectomy on the gross body activity of rats. *Endocrinology*, 1933, **17**, 445-450.
 77. RICHTER, C. P. Cyclical phenomena produced in rats by section of the pituitary stalk and their possible relation to pseudo-pregnancy. *Amer. J. Physiol.*, 1933, **106**, 80-89.
 78. RICHTER, C. P. Pregnancy urine given by mouth to gonadectomized rats: its effect on spontaneous activity and on the reproductive tract. *Amer. J. Physiol.*, 1934, **110**, 499-512.
 79. RICHTER, C. P. The spontaneous activity of adrenalectomized rats treated with replacement and other therapy. *Endocrinology*, 1936, **20**, 657-666.
 80. RICHTER, C. P., & BENJAMIN, J. A., JR. Ligation of the common bile duct in the rat. *Arch. Path.*, 1934, **18**, 817-826.

81. RICHTER, C. P., & ECKERT, J. F. The effect of hypophyseal injection and implants on the activity of hypophysectomized rats. *Endocrinology*, 1937, 21, 481-488.
82. RICHTER, C. P., & HARTMAN, C. G. The effect of injection of amiotin on the spontaneous activity of gonadectomized rats. *Amer. J. Physiol.*, 1934, 108, 136-143.
83. RICHTER, C. P., & HAWKES, C. D. Increased spontaneous activity and food intake produced in rats by removal of frontal poles of the brain. *J. Neurol. Psychiat.*, Chicago, 1939, 2, 231-242.
84. RICHTER, C. P., & HINES, M. Increased spontaneous activity produced in monkeys by brain lesions. *Brain*, 1938, 61, 1-16.
85. RICHTER, C. P., & SCHMIDT, E. C. H., JR. Behavior and anatomical changes produced in rats by pancreatectomy. *Endocrinology*, 1939, 25, 698-706.
86. RICHTER, C. P., & WANG, G. H. New apparatus for measuring the spontaneous motility of animals. *J. Lab. Clin. Med.*, 1926, 12, 289-292.
87. RICHTER, C. P., & WISLOCKI, G. B. Activity studies on castrated male and female rats with testicular grafts, in correlation with histological studies of the grafts. *Amer. J. Physiol.*, 1928, 76, 651-660.
88. RUCH, T. C., & SHENKIN, H. A. The relation of area 13 on the orbital surface of the frontal lobe to hyperactivity and hyperphagia in monkeys. *J. Neurophysiol.*, 1943, 6, 349-360.
89. RUNDQUIST, E. A. Inheritance of spontaneous activity in rats. *J. comp. Psychol.*, 1933, 16, 415-438.
90. RUNDQUIST, E. A., & BELLIS, C. J. Respiratory metabolism of active and inactive rats. *Amer. J. Physiol.*, 1933, 106, 670-675.
91. SCHULTE, J. W., TAINTER, M. L., & DILLE, J. M. Comparison of different types of central stimulation from analeptics. *Proc. Soc. exper. Biol. Med.*, 1939, 42, 242-248.
92. SEARLE, L. V., & BROWN, C. W. Effect of subcutaneous injections of benzedrine sulphate on the activity of white rats. *J. exper. Psychol.*, 1938, 22, 480-490.
93. SEARLE, L. V., & BROWN, C. W. Effect of variation in the dose of benzedrine sulphate on the activity of white rats. *J. exp. Psychol.*, 1938, 22, 555-563.
94. SHIRLEY, MARY. Studies in activity. I. The consistency of the revolving drum method of measuring the activity of the rat. *J. comp. Psychol.*, 1928, 8, 23-38.
95. SHIRLEY, MARY. Studies in activity. II. Activity rhythms, age and activity, activity after rest. *J. comp. Psychol.*, 1928, 8, 159-186.
96. SHIRLEY, MARY. Studies in activity. IV. The relation of activity to maze learning and brain weight. *J. comp. Psychol.*, 1928, 8, 187-195.
97. SHIRLEY, MARY. Spontaneous activity. *Psychol. Bull.*, 1929, 26, 341-365.
98. SIEGEL, P. S. A simple electronic device for the measurement of the gross bodily activity of small animals. *J. Psychol.*, 1946, 21, 227-236.
99. SIEGEL, P. S. Activity level as a function of physically enforced inaction. *J. Psychol.*, 1946, 21, 285-291.
100. SKINNER, B. F. The measurement of "spontaneous activity." *J. gen. Psychol.*, 1933, 9, 3-24.
101. SLONAKER, J. R. The normal activity of the white rat of different ages. *J. comp. Neurol. Psychol.*, 1907, 17, 342-359.
102. SLONAKER, J. R. Description of apparatus for recording the activities of small mammals. *Anat. Rec.*, 1908, 2, 116-122.
103. SLONAKER, J. R. Analysis of daily

- activity of the albino rat. *Amer. J. Physiol.*, 1925, **73**, 485-503.
104. SLONAKER, J. R. Pseudopregnancy in the albino rat. *Amer. J. Physiol.*, 1929, **89**, 406-416.
 105. SLONAKER, J. R. Effect of different per cents of protein in the diet. II. Spontaneous activity. *Amer. J. Physiol.*, 1931, **96**, 557-561.
 106. SLONAKER, J. R. Sex drive in rats. *Amer. J. Physiol.*, 1935, **112**, 176-181.
 107. SMITH, E. A., & CONGER, R. M. Spontaneous activity in relation to diet in the albino rat. *Amer. J. Physiol.*, 1944, **142**, 663-665.
 108. SMITH, P. K., & SMITH, A. H. The effect of a diet low in inorganic constituents on the voluntary activity of the albino rat. *Abstr. Proc. Amer. physiol. Soc.*, 1934.
 109. SMITH, K. U. An accurate method of recording activity in animals. *J. gen. Psychol.*, 1942, **27**, 355-358.
 110. STEWART, C. C. Variations in daily activity produced by alcohol and by changes in barometric pressure and diet, with a description of recording methods. *Amer. J. Physiol.*, 1898, **1**, 40-56.
 111. STIER, T. J. B. "Spontaneous activity" of mice. *J. gen. Psychol.*, 1930, **4**, 67-101.
 112. STONE, C. P. Effects of electro-convulsive shocks on daily activity of albino rats in revolving drums. *Proc. Soc. exp. Biol.*, N. Y., 1946, **61**, 150-151.
 113. STONE, C. P., & BARKER, R. B. Spontaneous activity, direct and indirect measures of sexual drives in adult male rats. *Proc. Soc. exp. Biol.*, N. Y., 1934, **32**, 195-199.
 114. SZYMANSKI, J. S. Eine methode zur Einversuchung der Ruhe und Aktivitätsperioden bei Tieren. *Arch. f. d. ges. Physiol.*, 1914, **158**, 343-385.
 115. SZYMANSKI, J. S. Aktivität und Ruhe bei Tieren und Menschen. *Z. allg. Physiol.*, 1920, **18**, 105-162.
 116. TAINTER, M. L. The effects of certain analeptic drugs on spontaneous running activity of the white rat. *J. comp. Psychol.*, 1947, **36**, 143-155.
 117. WALD, G., & JACKSON, B. Activity and nutritional deprivation. *Proc. nat. Acad. Sci.*, Wash., 1944, **30**, 255-263.
 118. WANG, G. H. The relation between spontaneous activity and estrous cycle in the white rat. *Comp. Psychol. Monogr.*, 1923, **2**, 1-27.
 119. WELSH, J. H. Diurnal rhythms. *Quart. Rev. Biol.*, 1938, **13**, 123-139.
 120. WERTHESEN, N. T. The significance of sub-normal respiratory quotient values induced by controlled feeding of the rat. *Amer. J. Physiol.*, 1937, **120**, 458-465.
 121. WILBUR, K. M. A method for the measurement of activity of small animals. *Science*, 1936, **84**, 274.
 122. YOUNG, W. C., & FISH, W. R. The ovarian hormones and spontaneous running activity in the female rat. *Endocrinology*, 1945, **36**, 181-189.
 123. ZIEVE, L. Effects of benzedrine on activity. *Psychol. Rec.*, 1937, **1**, 393-396.
 124. ZIEGLER, L. H., & KNUDSON, A. Activity after recovery from rickets. *J. comp. Psychol.*, 1936, **22**, 199-217.

SAMPLING IN THE REVISION OF THE STANFORD-BINET SCALE

ELI S. MARKS

National Office of Vital Statistics

In another paper (4) the writer attempts to point out the biases which may arise through types of sampling procedure quite common in psychological research. The present analysis is devoted to another effect of sampling methods commonly used in psychology—namely, the substantial increase in sampling error which results when “cluster” methods of sampling are used. It should be noted that this is not a criticism of the cluster type of sampling. Cluster sampling is an extremely valuable device and makes feasible many studies which would otherwise be completely impossible. However, the use of cluster techniques implies substantial modifications in our formulae for sampling error and psychologists are, in general, not familiar with these modifications. Unfortunately, much of the important work in the field appears in sources which are relatively inaccessible to psychologists. Ignoring the effects of cluster sampling on measures of sampling error has undoubtedly resulted in attaching importance to results which are statistically insignificant. In the testing field, failure to allow for cluster sampling has probably caused us to attach a measure of precision to our results considerably in excess of that warranted by sound statistical techniques.

Cluster sampling almost always involves an increase in sampling error as compared with unrestricted random sampling of the same number of cases. It is, of course, possible to obtain a lower sampling error with cluster sampling than with unrestricted random sampling if we make up our clusters for this purpose. However, the main reason for the use of cluster sampling is to permit the sampling of previously existing groups (the clusters) and, in most cases, the use of a previously existing grouping of the population involves a positive intraclass correlation of the variable studied, i.e., our existing groups are almost always more homogeneous internally than groups of the same size made up by random selection of individuals from the population. It is the existence of positive intraclass correlation which cuts down the amount of independent information available from a cluster sample of a specified size and occasions the substantial increase in sampling error usually associated with this sampling method. The present analysis is designed to emphasize the substantial increase in sampling error which results from relatively small intraclass correlations. While this phenomenon is quite familiar to sampling statisticians, psychologists are rather generally

unaware of the marked disturbances of sampling error calculations and tests of significance introduced by the use of cluster sampling when a positive intraclass correlation exists.

Although methods resembling cluster sampling are quite common in psychological research, very few psychological studies have used sampling designs which permit us to determine the standard error of the mean or of other sample statistics. As a matter of fact, it is difficult to find a study where analysis of the sampling error formulae used is not complicated by the presence of a non-measurable design (one in which the sampling probabilities are unknown). Some of the difficulties in the use of non-measurable designs are explored in a paper by McNemar (2) which discusses accidental sampling and purposive sampling as well as such measurable designs as unrestricted random sampling and stratified sampling.

THE SAMPLING PLAN OF THE STANFORD-BINET

The writer has, therefore, not attempted to find a study with a measurable design, but has selected for analysis the sample used in the revision of the Stanford-Binet. This sample has been selected for analysis principally because the widespread use of the revised Stanford-Binet makes the problems involved in its standardization extremely important in spite of the lapse of a decade since the revision was completed. The revision of the Stanford-Binet is also a good example for our purposes because (a) it was an extensive project, involving a relatively large number of subjects and the expenditure of considerable amounts of time, effort and money and (b) the purposes of the sample were explicitly formulated and clearly stated by the authors of the revised Stanford-Binet.

The reader should bear in mind that the present analysis is on an "as if" basis. The Stanford-Binet sampling design does not yield statistics with measurable standard errors and no amount of statistical manipulation can overcome this defect. The cure lies not in statistical formulae but in more careful sampling techniques in future investigations. However, the use of measurable sampling designs in psychological research will almost inevitably mean cluster sampling of some sort since any other approach will be beyond the limited resources usually available to psychologists. Thus, an examination of cluster sampling, even on an "as if" basis, is extremely pertinent to the future of any psychological research which involves statistical techniques.

In my analysis I have relied entirely upon statements and data published in Terman and Merrill (6) and McNemar (3). Since the data

required for this analysis have not been published in full detail, I have been forced to use approximations at several points. Inquiry indicated that more detailed data could not be furnished without considerable expenditure of time and effort. Since the approximations used in this paper are satisfactory for purposes of illustration and since the sampling techniques used in the revision of the Stanford-Binet preclude a completely accurate determination of error even if the detailed data were available, this deficiency is not serious. In nearly every case, the effect of the approximation used has been to understate the sampling error.

In revising the Stanford-Binet, the major objective was to construct scales "so standardized for difficulty as to yield mean I.Q.'s of approximately 100 at all age levels" (Terman, in 3, p. 3). The authors of the revision realized that their success in this objective was dependent upon securing a measure of the distribution of test scores in the general population (or in a satisfactory sample of the population). The sample was restricted to "American born" subjects of the "white race" in the age range from $1\frac{1}{2}$ years to 18 years. Terman notes that "elaborate precautions were taken to make the sampling as representative of the entire population as circumstances permitted" (3, p. 6).

According to Terman and Merrill this was done by selecting "17 different communities in 11 states" (6, p. 12). They note that: "The selection of localities for the second year's testing was based upon certain considerations in regard to sampling which had resulted from a study of the socio-economic level of the first 1500 subjects." These considerations were what the authors viewed as an inadequate representation of the rural group and a difference between the occupational distribution of fathers of the cases tested and the occupational distribution of all employed U. S. males. In the second year's testing, therefore, the authors of the Stanford-Binet revision "took care to include several additional rural communities" (6, p. 14). Neither McNemar nor Terman and Merrill give further details on the method of selecting the communities. It seems evident that selection was not on the basis of random sampling (neither simple random sampling nor random sampling within strata). As a matter of fact the term "community" is not defined clearly enough to permit a rigorous statement of the primary sampling units used. Nevertheless, we can visualize our population as being composed of "communities" (undefined but definable), so that the entire population of the United States can be broken up into a fairly large number (probably over 3000) of communities.

Within each community different procedures were followed for cases in the three age groups— $1\frac{1}{2}$ to $5\frac{1}{2}$ inclusive; 6 to 14 inclusive; and 15 to 18 inclusive. These groups were sampled as follows:

1. *The group aged 6 to 14.* Schools of "average social status" were selected in each community (method of selecting schools not further specified) and within each school all of the children between the ages 6 to 14 who were within one month of a birthday were taken, regardless of grade placement (6, p. 15). This sampling procedure is, then, a subsampling of subclusters with a 100 per cent sample take within the subcluster.

2. *The group aged 15 to 18.* Subjects were selected so that "the advanced group would be as nearly as possible continuous with the intermediate, with no break between fourteen and fifteen years. The compulsory school age was taken into account, the general character of the population, and the type of secondary education that was offered. In each community the school census was consulted to determine the amount of elimination after age fourteen. We made certain that some of the twelve-, thirteen-, and fourteen-year-olds who had gone to high school were included, also some of the slow fifteen- and sixteen-year-olds who were still in intermediate school. A few cases who had graduated from high school were included and a few who had dropped out of school without completing high school. These out-of-school groups were sampled by choosing siblings of school children in numbers proportional to the amount of elimination at ages above fourteen." (Sampling in this group was actually a rough type of "quota" sampling.)

3. *The group aged $1\frac{1}{2}$ to $5\frac{1}{2}$.* This group was sampled in much the same manner as the out-of-school cases in the group aged 15 to 18. The authors "chose as far as possible younger sibs of the school groups." Children were secured by use of birth records, school census, school siblings, kindergartens, well baby clinics, day nurseries, nursery schools and "personal report." Use of the various sources differed from community to community. "Great care was exercised in the large population centers to include representative groups; if a school in a suburban district which had been chosen as average on the advice of superintendent and counselors seemed to include too large percentage of higher occupational groups it was offset by a tenement district center." The authors state that "in the smaller communities, from seventy-five to eighty percent of the pre-school child population of appropriate age was examined" (6). In the published tabulations results for children aged $1\frac{1}{2}$ were omitted and further references deal only with the sample of children two years of age or over.

It may be noted that the population sampled is limited to individuals within one month of a birthday (or half-year birthday for children under six). The population is also limited to American-born white persons and, in the age range six to 14, to children attending school. These limitations do not affect the propriety of generalization from a sample to the population so defined. The limitations may affect generalization from the sample to all native-born white persons aged two to 18. This is not, however, of primary concern in this paper. Limitations on generalization resulting from the use of sub-populations are discussed by the writer in another article (4). For our present purposes, it is sufficient to accept the population, as defined.

To summarize, the sampling plan of the Stanford-Binet revision

involved: (a) sampling of "communities" from the aggregate of all United States communities; (b) the subsampling of schools for children aged six to 14 and taking all children (in the population as defined above) in the selected schools; (c) the subsampling of other members of the defined population from the "community" without any intermediate subsampling of schools but with the use of a rough type of "quota" sampling.

BIASES AND VARIANCE IN THE STANFORD-BINET SAMPLING

The above is only an approximate statement since it is extremely hard to formulate exactly the sampling plan used. The method of selection at each stage of sampling has not been specified above. It seems likely, however, that the sampling error of the plan used is greater than the error which would be involved in random sampling of "communities" with equal probability of selection and no subsampling (i.e., a plan which would take all persons in the communities selected).

It is obvious that a sampling plan not involving subsampling will have a lower sampling error for the same number of clusters sampled than a plan which did involve subsampling. The assumption that the community sampling actually used involved a larger sampling error than random selection is not as clear cut. Actually the sampling resembled "purposive sampling" or "quota sampling" but it does not appear to conform even to the rather loose requirements of these two techniques.

In discussing purposive sampling Neyman (5) developed certain hypotheses which, if satisfied, would make the estimate commonly used in this method the "best linear estimate" (i.e., an unbiased linear estimate with variance less than that of any other linear estimate). Neyman notes that:

If these hypotheses are not satisfied, which I think is a rather general case, we are not able to appreciate the accuracy of the results obtained. Thus this is not what I should call a representative method. Of course it may sometimes give perfect results, but these will be due rather to the uncontrollable intuition of the investigator and good luck than to the method itself.

While the Stanford-Binet revision did not involve purposive sampling, Neyman's remarks are applicable to the sampling plan. Furthermore, there is internal evidence in the results of the Stanford-Binet revision which indicates that, in spite of the purposive attempt to secure a "representative" sample, the Stanford-Binet revision sample actually produced a larger sampling error than would have resulted from random sampling of clusters.

Table 3 below gives the number of cases from each of the communities included in the Stanford-Binet sample. It should be noted that 37 percent of the "urban" cases were drawn from San Francisco and 56 percent of the "suburban" cases came from two California communities. This means that 975 cases or 34 percent of the total sample were from California. In addition a disproportionately large number of the "rural" cases (41 percent) came from one community in Vermont. It would, of course, be possible to obtain clusterings in two states as marked as those shown by a random sampling of communities, but the probability of such an outcome is extremely small. It is almost certain that a random (or stratified random) sampling of communities would have given a better geographic distribution (and undoubtedly a lower sampling error) than was actually obtained. This fact is also pointed out by McNemar (3) who expresses "skepticism concerning the representativeness of these communities."

It should also be remembered that the authors of the Stanford-Binet revision felt very definitely that their results contained a substantial bias. As noted above, the primary objective of the revision was to obtain a scale giving average I.Q.'s of 100 for each chronological age group. Terman and Merrill (6, p. 23) note that the mean I.Q.'s for their age groups run "slightly above 100" and state that this "is the result of intentional adjustment to allow for the somewhat inadequate sampling of subjects in the lower occupational classes." McNemar (3, p. 20) states:

The fact that the means in Tables 1 and 2 are above 100 should not lead the reader to the erroneous conclusion that the average I.Q. for the population now exceeds 100. The excess here observed is in the proper direction to allow for known bias in our age samplings. When an adjustment is made for bias in occupational status, the age means approach nearer 100, and a further adjustment for inadequate rural representation would tend to bring the values still closer to 100.

Table 6 on p. 36 of Terman and Merrill (6) gives average I.Q.'s for each age group "adjusted for 1930 Census frequencies of Occupational groupings." These averages still show substantial bias, all means except those for ages 4 and $5\frac{1}{2}$ being over 100 and seven age groups having average I.Q.'s over 103. The effect of rural-urban biasing influences is not presented.

Since the method of correcting for bias is not stated, the effect of these corrections on the mean square errors of the sample results cannot be determined. It is probably not possible to make this determination in any event since the presence or absence of biases in occupational or

rural-urban distributions cannot by themselves tell us whether an I.Q. distribution is biased or unbiased and correcting for rural-urban or occupational biases may have very little effect (or even an unfavorable effect) upon I.Q. biases.

In any event, the original sample means of the Stanford-Binet revision contain substantial biases if the true population means are 100. These are shown by the figures in Table 1.*

TABLE 1
AVERAGE I.Q.'s BY AGE GROUPS FOR THE STANFORD-BINET REVISION SAMPLE

	<i>Age Groups</i>			<i>All Cases</i>
	<i>2½-5½</i>	<i>6-13</i>	<i>14-18</i>	
From L—Mean	106.58	103.22	103.03	104.00
Form M—Mean	106.42	103.96	103.32	104.43
Number of Cases	728	1623	619	2970

In view of the probable biases and the considerations with regard to the sampling method presented above, it is not at all unreasonable to assume that the Stanford-Binet revision sampling involved a larger standard error of the mean than would random selection of communities with equal probability. Even if this is not the case, the subsampling involved should account for an increase in sampling error over a design in which there was no subsampling.

On the basis of the above discussion, the standard error of random selection of communities with equal probability and no subsampling gives us minimum values for the standard errors of the Stanford-Binet sample means. To estimate these errors, we shall assume that the number of cases actually sampled in each community was the total eligible population in that community. (As noted above, assuming that the community population was larger than the number sampled would lead to a larger estimate of the standard error.)

* The data are from McNemar (3) Tables 1 and 2. There are minor differences between the results presented by McNemar and those presented by Terman and Merrill (apparently due to inclusion of some subjects in some of the distributions and their omission in other distributions). The differences are minor and do not affect the present analysis.

The data in this and in the two subsequent tables are reproduced with the permission of Houghton Mifflin Co., the publishers of McNemar's *The Revision of the Stanford-Binet Scale*.

THE STANDARD ERROR FOR CLUSTER SAMPLING

The standard error for the type of sampling described (i.e. "cluster sampling") is given by:

$$\sigma_{\bar{x}}^2 = \frac{M - m}{(M - 1)m} \frac{\sum_i^M N_i^2 (\bar{x}_i - \bar{x})^2}{M \bar{N}^2} \quad [1]$$

Or, when we estimate $\sigma_{\bar{x}}$ from the sample, the estimated standard error is given by:

$$s_{\bar{x}}^2 = \frac{M - m}{Mm} \frac{\sum_i^m N_i^2 (\bar{x}_i - \bar{x}')^2}{(m - 1)(\bar{N}')^2} \quad [2]$$

where M = the total number of clusters (communities) in the population
 m = the number of communities sampled
 N_i = the number of individuals (eligible for the population) in the i -th cluster
 $\bar{x}_i$ = the mean I.Q. for the N_i individuals in the i -th cluster

$$\bar{x} = \frac{\sum_i^M N_i \bar{x}_i}{\sum_i^M N_i} = \text{the mean I.Q. of the population. (The aim of the sample is to estimate } \bar{x}.)$$

$$\bar{N} = \frac{\sum_i^M N_i}{M} = \text{the average number of individuals per cluster in the population.}$$

$$\bar{x}' = \frac{\sum_i^m N_i \bar{x}_i}{\sum_i^m N_i} = \text{the mean I.Q. of the sample. (We are using this as our estimate of } \bar{x} \text{ and } s_{\bar{x}} \text{ is the estimated standard error of this sample mean.)}$$

$$\bar{N}' = \frac{\sum_i^m N_i}{m} = \text{the average number of individuals per cluster in the sample.}$$

To determine $s_{\bar{x}}$ exactly we need to know M , the number of communities (clusters) in the population. While M is not known with any precision, we can be quite certain that it is large and that it is much

larger than m (at least 100 times as great). Consequently, we can, without appreciable error, take $M-m/M$ equal to 1. With this substitution the square of the standard error is approximately equal to:

$$s_{\bar{x}}^2 = \frac{\sum_i^m N_i^2 (\bar{x}_i - \bar{x}')^2}{m(m-1)(\bar{N}')^2} = \frac{m}{m-1} \frac{\sum_i^m N_i^2 (\bar{x}_i - \bar{x}')^2}{\left(\sum_i^m N_i\right)^2} \quad [3]$$

All the data required for Equation [3] can be obtained from the sample. Unfortunately, not all of the sample data are available in published form. Since we shall have to rely on published data, some further approximations (described below) are necessary. The approximations also act to reduce our estimate of the standard error.

McNemar (3) gives, as Table 9, information on the average I.Q.'s for children in "urban," "suburban" and "rural" communities by age groups. This table, plus data for the entire group in the age range 2 to 18, is presented in Table 2. The data for the entire group were calculated from the information given for the three age groups.

TABLE 2
I.Q. DATA FOR URBAN, SUBURBAN AND RURAL CHILDREN*

	Urban	Suburban	Rural	Total
<i>2-5½ Year-Olds</i>				
Number	354	158	144	656
Mean	106.3	105.0	100.6	104.7
S.D.	15.7	16.1	15.4	15.9
<i>6-14 Year-Olds</i>				
Number	864	537	422	1823
Mean	105.8	104.5	95.4	103.0
S.D.	14.7	16.8	15.5	16.1
<i>15-18 Year-Olds</i>				
Number	204	112	103	419
Mean	107.9	106.9	95.7	104.6
S.D.	16.5	15.7	15.9	16.9
<i>All Ages (2-18)</i>				
Number	1422	807	669	2898
Mean	106.2	104.9	96.6	103.6
S.D.	15.2	16.5	15.7	16.2

* Denver 2- to 5½-year-olds are excluded.

To determine $s_{\bar{x}}$ we shall take for our values of $\bar{x}_i$ (the mean I.Q. in each community): (a) the average I.Q. for urban children for each of

the communities classified as "urban by McNemar; (b) the average I.Q. for suburban children for each of the communities classified as "suburban" and (c) the average I.Q. for rural children for each of the communities classified as "rural." This approximation ignores all variations between communities within the urban, suburban and rural groups of communities. As a result the values of s_x , which we obtain should be equal to or less than the values which would be obtained if we knew the means of each of the sampled communities.*

There are some uncertainties in the published data concerning the values of m and N_i . As noted above, Terman and Merrill (6) state that 17 communities were sampled in 11 states. This would give $m=17$. However on pp. 36-37, McNemar (3) lists the communities sampled and the number of subjects in each community. McNemar lists 7 urban communities. He also lists 3 suburban communities and states that, in the suburban group, there were "four small communities just out of Kansas City in Johnson County Kansas, with 199 cases drawn from Westwood View, Hickory Grove, Roseland, and Shawnee Mission schools." For the rural communities, McNemar states:

The samplings from rural communities include 85 from Mount Washington School, Bullitt County, and Liberty School, Oldham County, Kentucky. A total of 152 were drawn from the following districts of Indiana: Prather School, Charlestown schools and Morgan Township School in Harrison County and Galena School in Floyd County. A farming region at Bloomington, Minnesota, supplied 92 cases; the farming and small village community of Randolph, Vermont, provided 275; and 65 subjects were secured in the vicinity of Atlee, Virginia. We have already expressed some skepticism concerning the representativeness of these communities.

From this statement, it is difficult to determine the exact number of "rural communities" involved. At a minimum, there appear to be 8 (assuming that schools in different counties represent different communities). If we also consider the four schools in the "suburban" part of Johnson County, Kansas, to be one community, McNemar's listing gives a count of 19 communities vs. Terman and Merrill's 17. The difference appears to be one in the definition of community. In terms of independently selected areas, Terman and Merrill's "17 communities" is probably more nearly correct. However, the data in Table 2 are based on McNemar's classification. It appears desirable to adopt a compromise, counting as communities the cities and towns listed by McNemar

* This statement cannot be made absolutely since, under certain circumstances, it may be incorrect. However, it is a fairly safe statement since the circumstances which would give a higher standard error through substituting group averages for individual averages are extremely unusual.

plus any schools in separate counties. This is the same basis we used in getting the count of 19 communities mentioned above. Since the count of independently sampled communities is probably Terman and Merrill's figure of 17, this handling of the problem operates in the same direction as the other approximations previously made.

The difficulty in determining N_i occurs in the cases where McNemar gives one figure for the number sampled in two different counties (e.g.,

TABLE 3
NUMBER OF CASES SAMPLED IN EACH COMMUNITY AND
ESTIMATED AGE DISTRIBUTION

<i>Communities</i>	<i>2-5½ Year-Olds</i>	<i>6-14 Year-Olds</i>	<i>15-18 Year-Olds</i>	<i>All Ages (2-18)</i>
<i>Urban</i>				
1. Denver, Col.	28	67	16	111
2. Minneapolis, Minn.	46	111	26	183
3. New York, N. Y.	12	29	7	48
4. Reno, Nev.	28	68	16	112
5. Richmond, Va.	46	114	27	187
6. San Antonio, Texas	63	155	36	254
7. San Francisco, Calif.	131	320	76	527
<i>Suburban</i>				
8. White Plains, N. Y.	31	107	22	160
9. Redwood City, Calif.	26	89	19	134
10. Los Gatos, Calif.	62	209	43	314
11. Johnson County, Kan.	39	132	28	199
<i>Rural</i>				
12. Bullitt County, Ky.	9	27	7	43
13. Oldham County, Ky.	9	27	6	42
14. Clark County, Ind.	11	32	8	51
15. Harrison County, Ind.	11	32	8	51
16. Floyd County, Ind.	11	31	8	50
17. Bloomington, Minn.	20	58	14	92
18. Randolph, Vt.	59	174	42	275
19. Atlee, Va.	14	41	10	65

85 cases from Bullitt County, Kentucky and Oldham County, Kentucky). These cases can be handled by distributing the cases equally among the counties involved. This adjustment also operates to reduce the estimated standard error. A further approximation is necessary to get standard errors for the means of each of the three age groups in Table 2. McNemar gives only the total number of cases in each community and does not give the distribution of these cases among the age groups. To estimate the standard errors for the separate age groups, the number

of cases for each of the communities was distributed by age proportionately to the age distribution in the class (urban, suburban or rural) in which the community falls. The number of cases in each community shown by McNemar and the calculated distribution of these cases by age groups is shown in Table 3. This adjustment affects only the estimates of the standard errors of the age group averages and not the standard error for the entire group aged 2 to 18.

COMPARISON OF CLUSTER SAMPLING ERROR WITH UNRESTRICTED RANDOM SAMPLING ERROR

With all the adjustments reducing the standard error which have been made, it may seem surprising that we have any error left. However, a fairly substantial amount of sampling error remains. Table 4 shows the standard errors of the mean I.Q. calculated as described above (using Equation 3) compared with the standard error obtained by the formula usually used in psychological research studies, i.e.:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{N'} \quad [4]$$

where

$$\sigma^2 = \frac{\sum_i^m \sum_j^{N_i} (x_{ij} - \bar{x}')^2}{N'} \quad [5]$$

and

$$N' = \sum_i^m N_i. \quad [6]$$

In Equations [4], [5] and [6], x_{ij} stands for the value (I.Q.) of the j th individual in the i th cluster (community) and the other symbols have the meanings previously defined. Equation [4] represents the standard error of the mean of a sample drawn by unrestricted random sampling from an infinite population (i.e. a sample drawn so that the probability of drawing any observation in the population is equal to and independent of the probability of drawing each of the other observations).

It will be seen from Table 4 that the absolute values of the standard errors calculated by Equation [3] are not large. There is a sampling error of only 1 per cent in the average I.Q. for the entire group of 2,898 cases. However, a very substantial difference exists between the stand-

ard error by Equation [3] and the standard error by Equation [4]. If we apply Equation [4] to determine the standard error of the mean of a cluster sample, it is obvious that we shall be very far from the correct value (in this case we would get an error which is less than one-third of the correct figure).

This fact is extremely important in applying tests of significance to differences of sample means. For example, suppose we took a sample

TABLE 4
ESTIMATED STANDARD ERRORS OF THE MEAN I.Q.'s FOR CLUSTER
SAMPLING AND UNRESTRICTED RANDOM SAMPLING

Age Groups	Standard Errors		Ratio of S.E. of Cluster Sampling to S.E. of Random Sampling
	Cluster Sampling	Unrestricted Random Sampling	
2-5½ years	.60	.62	.97
6-14 years	1.09	.38	2.89
15-18 years	1.35	.82	1.63
All Ages (2-18)	1.01	.30	3.36

of 900 children aged 2-18 (by a method which was actually random) from some city or other population subgroup. Assume that this sample gives us an average I.Q. of 105.7 on the revised Stanford-Binet and our sample has a standard deviation of 18, so that the standard error of the mean (using, quite properly, Equation [4]) is .60. Our group has a mean 2.1 points above the average of 103.6 for the Stanford-Binet revision sample shown in Table 2. We want to know whether this difference is significant. If we assume unrestricted random sampling of the Stanford-Binet revision sample, we would use .30 (see Table 4) as the standard error of the revision sample mean. This would give us .67 as the standard error of the difference of 2.1 and our difference would be 3.1 times its standard error. We would undoubtedly consider this a significant difference. Actually, the standard error of the mean of the revision sample is at least 1.01, which makes the standard error of the difference 1.17. The difference is actually only 1.8 times its standard error and can hardly be considered significant.

The sample used for the Stanford-Binet revision is not an extreme case of the error which can be made by applying formulae based on unrestricted random sampling to data obtained by cluster sampling. The sampling for the Stanford-Binet revision did involve testing individuals from several communities and the standard error for cluster sampling is

only 3 times the error for random sampling of the same number of individuals. Many studies use data from one or two groups (e.g.) elementary psychology classes in two neighboring colleges) to draw conclusions about the whole population (all college students or even all human beings). In this case the standard error obtained from Equation [3] may be 50 to 100 times greater than that obtained from Equation [4]. Use of the "correct" formula ("correct" if we have used a random process for drawing our groups) will make supposedly significant differences vanish more rapidly than a quart of ice cream at a children's party.

INTRACLASST CORRELATION

The reason for the difference between the standard error for unrestricted random sampling and that for cluster sampling is to be found in the fact that individuals are not sampled independently in cluster sampling. If we consider samples of equal size from the same population, the standard error of the mean in unrestricted random sampling is multiplied by approximately $(1 + \bar{N}\rho)$ when we use cluster sampling. Here $\bar{N}$ is the average size of our clusters and ρ is the intraclass correlation (a measure of the extent to which individuals within a cluster resemble, or are "correlated" with, each other). The intraclass correlation usually ranges from 0 to +1 (although it can be negative). It can be seen that even very small values of the intraclass correlation (say, .01) can have a very substantial effect on the standard error of a mean in cluster sampling if the clusters are moderately large ($\bar{N} = 100$ or more). As a matter of fact, the estimated intraclass correlation for the entire sample (all individuals aged 2 to 18) used in the Stanford-Binet revision is only .08. A recent paper by Walsh (7) gives some of the probability considerations involved in tests of significance when intraclass correlation is present.

There is one feature of Table 4 which may arouse some interest. This is the fact that the estimated standard error (using Equation [3]) of the mean I.Q. is larger for the group aged 6-14 years than for the group aged 2-5½ years. This is, of course, contrary to what we would expect from the description given of the sampling process. To some extent this peculiarity results from our ignoring subsampling in calculating the standard errors. Consideration of subsampling variation would probably increase the standard errors somewhat and would probably increase the standard error more for the group aged 2-5½ years than for the group aged 6-14 years (since there are fewer of the younger children). As a matter of fact, inclusion of subsampling variation might

double the standard error for the mean I.Q. of the group aged 2-5½ but would probably not increase the standard error of the group aged 6-14 more than 10 per cent

Actually Table 4 shows a lower standard error of cluster sampling for the group aged 2-5½ years than for the group aged 6-14 years because there is less variation among the average I.Q.'s of the urban, suburban and rural children for the younger group. This fact may be due to some basic relation between I.Q. variability and age. For example, McNemar (3) gives a table for adjusting I.Q.'s for differing standard deviation of the I.Q. at various ages. He bases this table on the differences actually found in the sample.

Another explanation of the differences in variability between age groups is to be found in the selective nature of the sampling for the Stanford-Binet revision. Selective sampling seems to have been particularly important in the pre-school group. In another article (4), the present writer points out some effects of selective sampling on group means and also notes that selective sampling will usually affect the standard deviation also. It would be very unwise to hypothesize about the difference between age groups shown in Table 4 unless we had much more information about the sampling probabilities.

This article has used the Stanford-Binet only as an illustration of the dangers of ignoring the intraclass correlation when we are dealing with a cluster type of sampling. In view of the qualifications placed on our analysis, it is not possible to draw any conclusions about the reliability or unreliability of the revised Stanford-Binet as a measuring instrument. There may be good reasons for supposing that the precision of the revised Stanford-Binet is considerably less than many of its users assume. From the sampling standpoint, the sample design used in the revision of the Stanford-Binet was a non-measurable design and there is no way of telling how "bad" or "good" the results were. It has been suggested that the sampling errors shown in Table 4 are probably minimum figures. However, the results do offer a possibility of improving the sampling design in the event that the Stanford-Binet is revised again in the future. An error of 1 I.Q. point in the average I.Q. may not be too serious. If this is the case, the biases in the Stanford-Binet average I.Q.'s could probably be removed by using a sound sample design without any need for an increase in either the number of communities covered or the number of subjects tested. If greater accuracy than a mean correct within 1 per cent is considered necessary or desirable, this could probably be achieved by increasing the number of communities sampled without increasing to any great extent the total

number of subjects tested. As a matter of fact, increasing the number of subjects tested would probably add very little to the accuracy of the final results (at least for the age group 6-14 years). The standard error of a mean in cluster sampling decreases (approximately) in proportion to the square root of the number of clusters sampled. The standard error shown in Table 4 for unrestricted random sampling is .3 of an I.Q. point. The standard error for cluster sampling is 3.36 times this value. Therefore, to get a standard error of .3 using cluster sampling, we would need about 11 times as many communities or about 200 communities. This estimate of the number of communities required is, of necessity, unreliable, since we were forced to estimate our standard errors from a sampling plan which is actually non-measurable, and measuring the non-measurable puts an obvious strain on epistemology.

In designing a sampling plan for a revision of the Stanford-Binet recent developments in sampling theory and practice can be used to increase accuracy without increase in survey costs. The reader's attention is directed particularly to the work of Hansen and Hurwitz (1) in this field. Using the techniques developed by Hansen and Hurwitz, persons revising the Stanford-Binet would probably get satisfactory precision from a well-designed sample of 25 to 100 communities with only a very small increase (if any) in the total number of cases tested.

SUMMARY

This article stresses the dangers of ignoring the intraclass correlation of the population when "cluster" methods of sampling are used. The increase in sampling error resulting from cluster sampling is demonstrated by an analysis of the results of the sample used in the revision of the Stanford-Binet. This sample actually yields "non-measurable" results, i.e. results which do not permit determination of the standard error of the sample mean. However, it is estimated that the standard error of the average I.Q. of this sample is at least 3 times the error which would be calculated by the use of the formula for unrestricted random sampling from an infinite population. The latter formula is the one familiar to psychologists and the one usually used by them regardless of the type of sampling involved. The illustration indicates that very substantial errors may result from this practice and that many results will be considered statistically significant where such a conclusion is entirely unwarranted.

BIBLIOGRAPHY

1. HANSEN, M. H., & HURWITZ, W. N. On the theory of sampling from finite populations. *Ann. math. Statist.*, 1943, 14, 333-362.
2. McNEMAR, QUINN. Sampling in psychological research. *Psychol. Bull.*, 1940, 37, 331-365.
3. McNEMAR, QUINN. *The revision of the Stanford-Binet Scale*. Boston: Houghton Mifflin, 1942.
4. MARKS, ELI S. Selective sampling in psychological research. *Psychol. Bull.*, 1947, 44, 267-275.
5. NEYMAN, JERZY. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *J. roy. statist. Soc.*, n. s., 1934, 97, 558-606.
6. TERMAN, L. M., & MERRILL, MAUD A. *Measuring intelligence*. Boston: Houghton Mifflin, 1937.
7. WALSH, J. E. Concerning the effect of intraclass correlation on certain significance tests. *Ann. math. Statist.*, 1947, 18, 88-96.

APPENDIX

Although the formula for the standard error of the mean for cluster sampling is not new, psychologists are generally unfamiliar with it. The derivation of this formula is, therefore, presented below. The z transformation will be found useful in deriving standard errors for more complicated designs (e.g. designs using stratification, subsampling, differential sampling probabilities, etc.).

Equation [2] gives the mean square error (square of the standard error) of the mean of a cluster sample as:

$$s_{\bar{x}}^2 = \frac{M - m}{Mm} \frac{\sum_i^m N_i^2 (\bar{x}_i - \bar{x}')^2}{(m - 1)(\bar{N}')^2}.$$

The mean of the sample is:

$$\bar{x}' = \frac{\sum_i^m N_i \bar{x}_i}{\sum_i^m N_i}.$$

M , m , N_i , $\bar{x}_i$, $\bar{x}'$, $\bar{N}'$, $\bar{x}$ and $\bar{N}$ are defined on p. 420. It is also convenient to define:

x_{ij} = value for j th individual in i th cluster

$x_i = \sum_j^{N_i} x_{ij}$ = sum of the values for all individuals in i th cluster.

From their definitions, it can be seen that:

$$\bar{x}_i = \frac{x_i}{N_i}$$

and, therefore

$$\bar{x}' = \frac{\sum_i^m x_i}{\sum_i^m N_i}.$$

$\bar{x}'$ can be treated as a ratio of two linear functions of the sample observations, namely:

$$f(x) = \frac{M}{m} \sum_i^m x_i = \frac{M}{m} \sum_i^m \sum_j^{N_i} x_{ij}$$

$$f(N) = \frac{M}{m} \sum_i^m N_i = \frac{M}{m} \sum_i^m \sum_j^{N_i} 1.$$

In deriving $s_{\bar{x}'}^2$, it will be useful to prove the following theorem:

Theorem: If we have a sample estimate:

$$r' = \frac{f(x)}{f(y)}$$

where $f(x)$ and $f(y)$ are linear functions of the sample observations x_h and $y_h (h=1, 2 \dots n)$ and if:

$$x = Ef(x), \quad y = Ef(y), \quad z_h = \frac{x_h}{x} - \frac{y_h}{y}$$

then:

$$\sigma_{r'}^2 = \left(\frac{x}{y} - r \right)^2 + \left(\frac{x}{y} \right)^2 \sigma_{f(x)}^2$$

where $\sigma_{r'}^2$ is the mean square error of r' and r is the population parameter of which r' is an estimate.

Proof: When we have a sample estimate $r' = f(x)/f(y)$, the mean square error of r' can be found by: (a) expanding r' as a Taylor series around x and y (the expected values of $f(x)$ and $f(y)$); (b) subtracting r (the true value of r' for the entire population) from both sides of the equation; (c) squaring both sides of the resulting equation and (d) taking the expected value of the resultant. If we ignore, in our Taylor series, terms involving partial derivatives higher than the first, the result of this operation will be:

$$\sigma_{r'}^2 = E(r' - r)^2$$

$$= \left(\frac{x}{y} - r \right)^2 + \left(\frac{x}{y} \right)^2 \left(\frac{\sigma_{f(x)}^2}{x^2} + \frac{\sigma_{f(y)}^2}{y^2} - 2 \frac{\sigma_{f(x)f(y)}}{xy} \right). \quad [7]$$

If we let

$$z_h = \frac{x_h}{x} - \frac{y_h}{y}$$

and if f is a linear function, then:

$$f(z) = \frac{f(x)}{x} - \frac{f(y)}{y} \quad [8]$$

and

$$Ef(z) = \frac{Ef(x)}{x} - \frac{Ef(y)}{y} = \frac{x}{x} - \frac{y}{y} = 0 \quad [9]$$

$$\sigma_{f(z)}^2 = \frac{\sigma_{f(x)}^2}{x^2} + \frac{\sigma_{f(y)}^2}{y^2} - \frac{2\sigma_{f(x)f(y)}}{xy} \quad [10]$$

Therefore:

$$\sigma_{r'}^2 = \left(\frac{x}{y} - r\right)^2 + \left(\frac{x}{y}\right)^2 \sigma_{f(z)}^2 \quad [11]$$

The above theorem can be applied to derive the mean square error of $\bar{x}'$, as follows:

$$\bar{x}' = \frac{f(x)}{f(N)}$$

where

$$f(x) = \frac{M}{m} \sum_i^m x_i \quad \text{or} \quad f(x) = \frac{M}{m} \sum_i^m \sum_j^{N_i} x_{ij}$$

$$f(N) = \frac{M}{m} \sum_i^m N_i \quad f(N) = \frac{M}{m} \sum_i^m \sum_j^{N_i} N_{ij} \quad \text{where } N_{ij} \equiv 1$$

and

$$Ef(x) = \sum_i^M x_i = x$$

$$Ef(N) = \sum_i^M N_i = N.$$

Let

$$z_i = \frac{x_i}{x} - \frac{N_i}{N} \quad \text{or}^* \quad z_{ij} = \frac{x_{ij}}{x} - \frac{1}{N}$$

$$f(z) = \frac{M}{m} \sum_i^m z_i = \frac{\frac{M}{m} \sum_i^m x_i}{x} - \frac{\frac{M}{m} \sum_i^m N_i}{N}.$$

By Equation [11]:

$$\sigma_{\bar{x}'}^2 = \left(\frac{x}{N} - \bar{x} \right)^2 + \left(\frac{x}{N} \right)^2 \sigma_{f(z)}^2 \quad [12]$$

where $\bar{x}$ is the population parameter estimated by $\bar{x}'$ and is:

$$\bar{x} = \frac{\sum_i^M x_i}{\sum_i^M N_i} = \frac{x}{N}.$$

Therefore:

$$\sigma_{\bar{x}'}^2 = \left(\frac{x}{N} \right)^2 \sigma_{f(z)}^2. \quad [13]$$

Since $f(z)$ is M/m times a sum of the sample values z_i :

$$\sigma_{f(z)}^2 = \frac{M^2(M-m)}{(M-1)m} \frac{\sum_i^M (z_i - \bar{z})^2}{M} \quad [14]$$

where

$$\bar{z} = \frac{\sum_i^M z_i}{M} = Ef(z) = 0. \quad [15]$$

An unbiased estimate of $\sigma_{f(z)}^2$ from the sample is:

$$s_{f(z)}^2 = \frac{M^2(M-m)}{Mm} \frac{\sum_i^m (z_i - \bar{z}')^2}{m-1} \quad (16)$$

* The result is the same whether the z transformation is applied to the cluster totals or the individual observations.

where

$$\bar{z}' = \frac{\sum_i^m z_i}{m} = \frac{\sum_i^m x_i}{mX} - \frac{\sum_i^m N_i}{mN}.$$

From Equations [13], [14], and [16] we have:

$$\sigma_{\bar{z}'}^2 = \left(\frac{x}{N}\right)^2 \frac{M^2(M-m)}{(M-1)m} \frac{\sum_i^M (z_i - \bar{z})^2}{M} \quad [17]$$

and

$$s_{\bar{z}'}^2 = \left(\frac{x}{N}\right)^2 \frac{M^2(M-m)}{Mm} \frac{\sum_i^m (z_i - \bar{z}')^2}{m-1}. \quad [18]$$

In Equation [17] we substitute the values:

$$z_i = \frac{x_i}{x} - \frac{N_i}{N}, \quad \bar{z} = 0, \quad \bar{N} = \frac{\sum_i^M N_i}{M}$$

and get:

$$\sigma_{\bar{z}'}^2 = \frac{(M-m)}{(M-1)m} \frac{\sum_i^M \left(x_i - N_i \frac{x}{N}\right)^2}{M\bar{N}^2} \quad [19]$$

or

$$\sigma_{\bar{z}'}^2 = \frac{M-m}{(M-1)m} \frac{\sum_i^M N_i^2 (\bar{x}_i - \bar{x})^2}{M\bar{N}^2}. \quad [20]$$

We make the same substitutions in Equation [18] and also substitute for $\bar{x}$ and $\bar{N}$ the sample estimates:

$$\bar{x}' = \frac{f(x)}{f(N)} = \frac{\sum_i^m x_i}{\sum_i^m N_i}$$

and

$$\bar{N}' = \frac{f(N)}{M} = \frac{\sum_i^m N_i}{m}.$$

This gives:

$$s_{\bar{x}'}^2 = \frac{M-m}{Mm} \frac{\sum_i^m N_i^2 (\bar{x}_i - \bar{x}')^2}{(m-1)(\bar{N}')^2} \quad [21]$$

or

$$s_{\bar{x}'}^2 = \frac{M-m}{M} \frac{m}{m-1} \frac{\sum_i^m N_i^2 (\bar{x}_i - \bar{x}')^2}{\left(\sum_i^m N_i\right)^2}. \quad [22]$$

In some cases, cluster sampling may introduce a substantial bias into the sample standard deviation (when the sample S.D. is used as an estimate of the population S.D.). This bias will be practically eliminated by use of the estimate:

$$s_s^2 = \sigma_s^2 + s_{\bar{x}'}^2 \quad [23]$$

where σ_s is the sample S.D. and s_s is an estimate of the population S.D. Equation [23] can also be used for estimating the population S.D. from a sample with unrestricted random sampling.

ILLUMINATION STANDARDS FOR EFFECTIVE AND EASY SEEING

MILES A. TINKER

University of Minnesota

The problem of artificial illumination is of primary importance in all inside working environments. To maintain healthful and efficient functioning of the eyes, it is necessary to provide adequate lighting. Unquestionably, proper illumination contributes much to comfort and efficiency in activities of daily life. Working under faulty illumination frequently results in eyestrain which tends to be accompanied by reflex functional disturbances of other organs.

During recent years a "lighting consciousness" has been forced upon a large portion of the population, particularly upon those who do considerable visual work under artificial light and upon those who must decide upon the illumination requirements of schools, offices, factories and other situations where visual work is to be performed. Although interest in lighting has been stimulated by popular articles, advertisements, and "educational pamphlets"—as well as by reports written by educators and medical men—the more fundamental information has appeared as experimental reports in scientific publications. The result of exposure to this material is a keen interest in illumination and a sincere desire on the part of the public for sound information concerning hygienic lighting. The natural tendency is to consult pamphlets on recommended practice when lighting specifications are needed for a particular situation. Frequently, the applied psychologist will be called upon to furnish advice on proper illumination. In many instances he will be asked to evaluate the materials presented in the recommended practices. Consequently, the applied psychologist should be informed concerning the adequacy of the data from which the lighting specifications in the recommendations are derived.

The first code on lighting was issued by the Illuminating Engineering Society in 1915. In the more recent publications, the codes are known as Recommended Practice of Home Lighting, of Office Lighting, etc. These pamphlets have been prepared by the Illuminating Engineering Society either alone or jointly with the American Institute of Architects, usually under the rules of procedure of the American Standards Association. Although the American Psychological Association has been in existence for over 50 years, and even though applied psychologists have been interested in the field and have been making experimental contributions to the hygiene of vision for over 40 years, neither psy-

chology nor psychologists are represented in the group specifying recommended practices. Furthermore, a large body of psychological literature has been ignored, either because the illuminating engineers were not familiar with it or because they chose not to use it. The result has been an emphasis upon the engineering aspects of lighting with inadequate attention to certain psychological factors. More recently there has been some attempt to consider more of the psychological factors. Perhaps because engineers lack a psychological background, interpretations are frequently erroneous. Probably the most satisfactory approach to hygienic lighting could be achieved by coordinating the contributions of engineers, physiologists, and psychologists.

Recent editions of recommended practices reveal an increased emphasis upon control of direct and reflected glare, brightness contrast, and the diffusion or distribution of light. The tendency to specify relatively very intense light for many visual tasks is prominent. The purpose of this paper is to present a critical examination of the specifications in the more recent editions of recommended practices and to scrutinize some of the data from which the recommendations were derived.

SPECTRAL QUALITY OF LIGHT

In general, spectral quality of light receives adequate treatment in recommended practices (35, 36, 37, 38). It is stated that with equal foot candles of illumination, variations in color quality of light found in common illuminants have little or no effect upon the visual discrimination involved. When color is to be discriminated, it should be viewed under as close an approximation of daylight as possible. Luckiesh (10) has a valuable discussion of light and color.

QUALITY OF LIGHTING

Recommendations (35, 36, 37, 38) concerning control of glare, diffusion, direction and distribution of light, light reflection value, and effects of finishes on ceilings and wall are ordinarily quite satisfactory. Visual discrimination is improved by moving the glare source away from the line of vision and by reducing the brightness of the light source and the amount of light emitted by the light source toward the eye. Brightness of luminaires should be low in value. High brightness contrasts within the field of vision should be avoided whether on the work surface or in other parts of the visual field. Proper diffusion of light helps to eliminate undesirable shadows. Purely local lighting, therefore, is unsatisfactory. Since the reflection factors of objects in the visual environment play an important role in illumination, the finish of ceilings, walls,

floors and furnishings is important. These surfaces should provide reflecting surfaces to help spread the light about the room. Furthermore, they should be such that undesirable brightness contrast does not occur within the field of vision. Shiny or glossy finishes should be avoided to prevent specular glare.

In the recommended practices, informative discussions on classification of lighting systems are usually included. Also illustrations of fixtures and installations are sometimes given. Some attention is given to daylight illumination and the need of coordinating artificial with daylight lighting.

INTENSITY OF ILLUMINATION

Intensity of illumination receives by far the greatest emphasis in specifications. With each revision of a lighting code prepared by illuminating engineers, the foot candle recommendations for a given situation rise. One may well question whether this trend has a scientific basis, or whether the consumer has been educated to accept the higher intensities. In 1934, Luckiesh and Moss (11) presented general recommendations which they considered to be very conservative. These are repeated with slight changes in Luckiesh's 1944 book (10). He adds that these are inadequate in many cases where hundreds and even thousands of foot candles of light are desirable. Examination of the recommended practices of lighting reveals that, for the most part, they are based upon researches done and interpretations made by Luckiesh and his co-workers, or upon researches inspired by them. Let us turn first, therefore, to these reports.

In *Light, Vision and Seeing*, Luckiesh (10), and in the *New Science of Seeing*, Luckiesh and Moss (11), make the following foot candle recommendations for common tasks of the work-world:

1. 100 foot candles or more are specified for severe and prolonged visual work. Examples include fine needle work, pen work, engraving and assembly, and discrimination of fine details involving low contrast.
2. 50 to 100 foot candles should be used for proof-reading, difficult reading, watch repairing, and average sewing.
3. 20 to 50 foot candles are listed for such visual tasks as clerical work, ordinary reading and average sewing on light goods.
4. 10 to 20 foot candles are proposed for ordinary reading and sewing on light goods when the task is not prolonged.
5. 5 to 10 foot candles are needed for visual work which is more or less interrupted or casual.
6. 1 to 5 foot candles are sufficient for perceiving large objects.

Luckiesh (10) states that these are minimum foot candle recommendations and that he considers them to be very conservative from the

viewpoint of ease of seeing. Furthermore these foot candles, according to Luckiesh and Moss (11), are far below the intensities of illumination which new knowledge indicates to be ideal.

These recommendations are derived from various sets of data which will be discussed in turn.

Preferences for light intensity. Luckiesh and Moss (11) cite data on preferences for light intensities to support their contentions that high intensities are necessary for adequate seeing. The mean choice was about 100 foot candles but the median was 50 foot candles when up to 1000 foot candles were available. Tinker's analysis (22) of light preference studies indicated that visual adaptation plays an important role in determining the preferences. In an experimental check, Tinker (26) found that when readers were adapted to 8 foot candles, the median choice for comfortable reading was about 12 foot candles. But when adapted to 52 foot candles, the median choice was 52 foot candles. It is obvious that the intensity of illumination to which the reader is adapted plays a dominant role in his illumination preference. The conclusion is, therefore, that preference for illumination intensity is not a satisfactory method for determining the intensity of light needed for efficient visual work.

Visual acuity. Luckiesh and Moss (11) and Luckiesh (10) list visual acuity as a basic factor in reading (and presumably in other visual work). It is true enough that visual discrimination does depend somewhat upon visual acuity. But is visual acuity an adequate criterion for prescribing appropriate lighting? Luckiesh and Moss (13) admit that in many tasks the criterion of visual acuity is relatively inappropriate, e.g. in tasks involving low contrasts. But they point out that for black test objects on a white background, visual acuity improves up to 100 foot candles. As a matter of fact, Lythgoe (15) has shown that under certain conditions of measurement, visual acuity improves up to and beyond 1000 foot candles. Inspection of the data reveal that the knee of the curve of improvement is at about 10 foot candles and that beyond about 20 foot candles the gains are slight. It must be kept in mind that in measuring visual acuity, one is dealing with threshold values. It is highly questionable whether the almost microscopic gains in visual acuity obtained under the high foot candles justify their application to visual tasks where supra-threshold visibility is involved as in most everyday situations. Furthermore, data reveal that the visual acuity curve is practically horizontal from 50 foot candles to the higher levels.

Luckiesh and Moss (11) and Luckiesh (10) cite data on visual acuity for 1, 10, and 100 foot candles only. If they really desired to find the foot candle level beyond which no *practical* gains in visual acuity occur, they should have investigated the range between 10 and 100 foot can-

dles. As shown in Tinker's reviews (29, 31), this criticism may be aimed at all the basic data presented by Luckiesh (10). In some instances (decrease in heart rate, decrease in convergence reserve of ocular muscles), data for only 1 and 100 foot candles are presented. This procedure is inexcusable in experiments designed to determine how much light intensity is needed for efficient visual work. It appears, then, that visual acuity data are of only slight use for prescribing illumination intensities for visual discrimination in supra-threshold tasks. If accepted, there is no justification for suggesting that more than 40 to 50 foot candles are necessary for adequate discrimination even for tasks that approach threshold discrimination.

Visibility measurements. Luckiesh (10) states that "After establishing a standard of visibility or desirable see-level to be attained if possible for all tasks, it is seen that specifications of light and lighting and other aids to seeing can be based upon visibility measurements." The measurements are to be made by the Luckiesh-Moss Visibility Meter. This is a device consisting of two identical circular gradients which are rotated before the eyes to alter the brightness contrast of the object whose visibility is to be measured. It, therefore, reduces the object to threshold visibility. It is the threshold which is measured. Three assumptions are made: (a) Two objects are equal in visibility when both are barely visible, (b) "Two objects are equally above threshold visibility when their visibility has been increased by the same increase" in size, brightness, brightness contrast or time, (c) "The visibility of an object, or degree of supra-threshold visibility, is proportional to the decrease in any one of the fundamental factors necessary to reduce the object to threshold visibility." These assumptions are considered to be *axiomatic* and arguments against them are considered to be *futile*. Nevertheless, since recommended standards are based upon visibility measurements to a large degree, it seems desirable to examine the matter further. Things are not axiomatic just because some one says they are.

Since visibility measurements are in terms of threshold values, they are analogous to visual acuity measurements. They are subject, therefore, to the same criticisms as visual acuity measurements as criteria for prescribing illumination standards.

Luckiesh (10) emphasizes foot candles for equal visibility in prescribing illumination intensities. For example, to make newspaper text matter equivalent in visibility to 8 point book type on white paper under 10 foot candles of light, it is necessary to use 30 foot candles. And to make the 1/64" divisions on a steel scale equal to this visibility level, 180 foot candles are needed. Are these levels of illumination intensity required for efficient and comfortable seeing? Luckiesh (10) assumes that this is a conservative standard. On his empirical scale, the 8 point type with 10 foot candles has 48 per cent maximum visibility. (Maxi-

viewpoint of ease of seeing. Furthermore these foot candles, according to Luckiesh and Moss (11), are far below the intensities of illumination which new knowledge indicates to be ideal.

These recommendations are derived from various sets of data which will be discussed in turn.

Preferences for light intensity. Luckiesh and Moss (11) cite data on preferences for light intensities to support their contentions that high intensities are necessary for adequate seeing. The mean choice was about 100 foot candles but the median was 50 foot candles when up to 1000 foot candles were available. Tinker's analysis (22) of light preference studies indicated that visual adaptation plays an important role in determining the preferences. In an experimental check, Tinker (26) found that when readers were adapted to 8 foot candles, the median choice for comfortable reading was about 12 foot candles. But when adapted to 52 foot candles, the median choice was 52 foot candles. It is obvious that the intensity of illumination to which the reader is adapted plays a dominant role in his illumination preference. The conclusion is, therefore, that preference for illumination intensity is not a satisfactory method for determining the intensity of light needed for efficient visual work.

Visual acuity. Luckiesh and Moss (11) and Luckiesh (10) list visual acuity as a basic factor in reading (and presumably in other visual work). It is true enough that visual discrimination does depend somewhat upon visual acuity. But is visual acuity an adequate criterion for prescribing appropriate lighting? Luckiesh and Moss (13) admit that in many tasks the criterion of visual acuity is relatively inappropriate, e.g. in tasks involving low contrasts. But they point out that for black test objects on a white background, visual acuity improves up to 100 foot candles. As a matter of fact, Lythgoe (15) has shown that under certain conditions of measurement, visual acuity improves up to and beyond 1000 foot candles. Inspection of the data reveal that the knee of the curve of improvement is at about 10 foot candles and that beyond about 20 foot candles the gains are slight. It must be kept in mind that in measuring visual acuity, one is dealing with threshold values. It is highly questionable whether the almost microscopic gains in visual acuity obtained under the high foot candles justify their application to visual tasks where supra-threshold visibility is involved as in most everyday situations. Furthermore, data reveal that the visual acuity curve is practically horizontal from 50 foot candles to the higher levels.

Luckiesh and Moss (11) and Luckiesh (10) cite data on visual acuity for 1, 10, and 100 foot candles only. If they really desired to find the foot candle level beyond which no *practical* gains in visual acuity occur, they should have investigated the range between 10 and 100 foot can-

dles. As shown in Tinker's reviews (29, 31), this criticism may be aimed at all the basic data presented by Luckiesh (10). In some instances (decrease in heart rate, decrease in convergence reserve of ocular muscles), data for only 1 and 100 foot candles are presented. This procedure is inexcusable in experiments designed to determine how much light intensity is needed for efficient visual work. It appears, then, that visual acuity data are of only slight use for prescribing illumination intensities for visual discrimination in supra-threshold tasks. If accepted, there is no justification for suggesting that more than 40 to 50 foot candles are necessary for adequate discrimination even for tasks that approach threshold discrimination.

Visibility measurements. Luckiesh (10) states that "After establishing a standard of visibility or desirable see-level to be attained if possible for all tasks, it is seen that specifications of light and lighting and the other aids to seeing can be based upon visibility measurements." The measurements are to be made by the Luckiesh-Moss Visibility Meter. This is a device consisting of two identical circular gradients which are rotated before the eyes to alter the brightness contrast of the object whose visibility is to be measured. It, therefore, reduces the object to whose visibility is to be measured. Three assumptions are made: (a) Two objects are equal in visibility when both are barely visible, (b) "Two objects are equally above threshold visibility when their visibility has been increased by the same increase" in size, brightness, brightness contrast or time, (c) "The visibility of an object, or degree of supra-threshold visibility, is proportional to the decrease in any one of the fundamental factors necessary to reduce the object to threshold visibility." These assumptions are considered to be *axiomatic* and arguments against them are considered to be *futile*. Nevertheless, since recommended standards are based upon visibility measurements to a large degree, it seems desirable to examine the matter further. Things are not axiomatic just because some one says they are.

Since visibility measurements are in terms of threshold values, they are analogous to visual acuity measurements. They are subject, therefore, to the same criticisms as visual acuity measurements as criteria for prescribing illumination standards.

Luckiesh (10) emphasizes foot candles for equal visibility in prescribing illumination intensities. For example, to make newspaper text matter equivalent in visibility to 8 point book type on white paper under 10 foot candles of light, it is necessary to use 30 foot candles. And to make the 1/64" divisions on a steel scale equal to this visibility level, 180 foot candles are needed. Are these levels of illumination intensity required for efficient and comfortable seeing? Luckiesh (10) assumes that this is a conservative standard. On his empirical scale, the 8 point type with 10 foot candles has 48 per cent maximum visibility. (Maxi-

imum visibility is obtained from a test-object whose critical detail has a visual size of 20 minutes; a critical detail of 1 minute is the smallest visible for persons with normal vision under 10 foot candles of light.) But no adequate experimental check is made for performance of these tasks under various levels of illumination. Tinker (27) found that the critical illumination level (the intensity beyond which no further change in reading performance occurs as the intensity is increased) for reading 7 point newspaper type to be approximately 7 foot candles. It is difficult to conceive the need of going above 20 foot candles to provide a margin of safety above the critical level. It is highly probable that an experimental check will reveal that other visual tasks, like discriminating the divisions on a steel scale, do not require the 180 foot candles indicated for efficient vision by the *computations* of Luckiesh. Related to this is the question of comfortable vision. Harrison (8), in discussing the difficulty of using high intensities because of the introduction of glare factors states "Visibility and comfort are two separate factors which do not always overlap completely."

No one will deny that visibility is an important factor in ease of seeing. But to prescribe standards in terms of scores derived from measurements made with the Visibility Meter is open to serious question. The basic data are threshold scores. While the derived scores may appear logical, supra-threshold seeing is not the same phenomenon as threshold seeing. Apparently, as illumination intensity is increased, one soon reaches a level of diminishing returns where further increase is of no practical importance or may introduce harmful factors from the viewpoint of easy and comfortable seeing.

Nervous muscular tension. Luckiesh and Moss (11, 12), place great stress upon the apparent decrease in nervous muscular tension during reading as the illumination intensity is increased from 1 to 10 to 100 foot candles. Tinker's (22) analysis of their data reveals that the method employed to present their results magnifies minute differences so that they appear large. Interpolation shows only gradual changes from 10 to 20 to 25 foot candles and very slight changes from there on to 100 foot candles. The conclusions that high foot candles are needed for ordinary reading is not valid. In a comparable situation, Tinker (25) found that for reading 10 point type, the critical intensity was about 3 foot candles. Below this level, rate of reading was retarded and fatigue increased, but for higher intensities there was no change. For people with normal vision, 10 to 15 foot candles should provide a satisfactory margin of safety for reading legible print.

Frequency of blinking. Another favorite criterion employed by Luckiesh and Moss (11, 12) and Luckiesh (10) as a basis for prescribing illumination intensities for visual work is frequency of blinking. The typical experiment is to measure the rate of involuntary blinking for the first and for the last five minutes for an hour's reading under 1, under

10 and under 100 foot candles of light. They note that the blink rate is greater under the 1 than under 10, and greater under 10 than under 100 foot candles. Therefore, it is concluded that relatively high intensities are desirable for reading. Even if these data are accepted as valid, we do not know where between 10 and 100 foot candles the curve of increased efficiency flattens out since intermediate intensity values were not studied. But there are several sources of information which suggest that blink rate is not a valid criterion of ease of seeing:

1. McFarland, Holway, and Hurvich (18), after a searching analysis of their own extensive experiments and of other studies, state: "A high blink-rate need mean neither an increase in fatigue nor an increase in difficulty of seeing." They conclude that "the rate of blinking can hardly be considered as a valid index of visual fatigue."

2. Tinker (32), in a study that has some bearing on the subject, found that frequency of blinking is an inadequate criterion of readability of print.

3. Bitterman (1), working with 3 and 91 foot candles of light, found that when subjects read for 40 minutes there was no significant difference in rate of blinking. In fact the frequency of blinking was slightly greater under the 91 foot candles. Incidentally, Bitterman also found no significant difference in blink rate for reading large type vs. small type. His results, therefore, indicate that rate of blinking cannot be employed as an index of ease of visual work. Further studies by Bitterman and Soloway (2, 3) showed that frequency of blinking is unrelated to duration of visual work or to the presence of a relatively intense glare source in the visual field. The reports of McNally (19) and MacPherson (16) also cast doubt upon the validity of blinking as an index of ease of seeing.

4. The statistical treatment employed by Luckiesh and Moss (11, 12, 14) upon their data is open to severe criticism. Tinker (28, 29) has questioned the appropriateness of the geometric mean which they employ in most comparisons. The same criticism is raised by Hoffman (9). In a searching analysis, Hoffman also severely criticizes the use of the percentage technique employed by Luckiesh and Moss for presenting data, and for basing conclusions on percentage differences rather than on raw score differences. Percentage scores are notoriously unreliable. Furthermore, if the raw scores are below 100 (as most of them are), percentages magnify the differences. When percentages are used, therefore, the observed differences may be largely an effect of the derivation. Insignificant raw score differences may seem large when put into percentages. For instance, a typical average of 30 blinks during 5 minutes of reading is increased 10 per cent by a change of 3 blinks. Hoffman further points out that work decrement may be a more important variable than illumination changes in the results of Luckiesh and Moss. In general, he found little support for the contention that relatively high intensities are needed for effective and easy seeing.

5. Eames (5) criticizes Luckiesh and Moss (14) for using relatively few subjects in their experiments (including blink rate studies) and for employing "test wise" subjects. As pointed out by Eames, "People who take tests repeatedly in a given field gradually learn what is expected of them" and are un-

intentionally influenced by this knowledge. Results obtained under such conditions cannot be representative of the reactions of the general population.

The accumulated evidence indicates that rate of blinking cannot be accepted as a criterion for specifying intensities of light for visual work.

Decrease in heart rate. Luckiesh (10) and Luckiesh and Moss (11, 12, 14) cite data on change of heart rate while reading for one hour under 1 foot candle and under 100 foot candles of light. No data are presented for intermediate levels of illumination. It is stated that heart rate decreased 10 per cent under the 1 foot candle and 2 per cent under 100 foot candles. The conclusion was that from the viewpoint of ease of seeing the 100 foot candle level is desirable. An experiment by McFarland, Knehr and Berens (17) was designed to check the findings obtained in Luckiesh's laboratory. The results led to the conclusion that "It is questionable whether reliable criteria for determining adequate levels of illumination for tasks such as reading during short periods of time (approximately 2 hours) can be obtained in terms of . . . heart rate" Another check experiment was carried out by Bitterman (1), who recorded heart rate while reading under 3 and under 91 foot candles of light. "The results do not support the conclusions of Luckiesh and Moss with respect to the value of heart rate" as an index "of the ease of visual work." In view of the above evidence we must reject heart rate as a criterion for prescribing illumination intensities for visual work.

Decrease in convergence reserve. Luckiesh and Moss (11, 12, 14) and Luckiesh (10) cite data on decrease in convergence reserve of ocular muscles after reading for one hour under 1 and under 100 foot candles of light. The decrease was less under the 100 foot candles. No data are given for the range between 10 and 100 foot candles. We do not know, therefore, whether the 100 foot candles is significantly better than such levels as 20 or 30 foot candles.

Visual adaptation. Throughout their writings, Luckiesh and Moss (10, 11, 12, 14) emphasize that the eyes evolved under daylight levels of illumination and suggest the desirability of competing with daylight by artificial means. They consistently ignore the fact that the eyes readily adapt to easy and effective seeing over a wide range of illumination intensities.

Summary on intensity of illumination. Examination of the data employed by Luckiesh and Moss as a basis for specifying foot candle levels for visual work reveals a general lack of validity of these results as criteria for ease of seeing. The data from visual acuity, muscular tension and visibility measurements are misinterpreted or misapplied. The blink technique and rate of heart beat must be rejected because of lack of confirmation by independent workers. Furthermore the methods of

statistical analysis employed are frequently at fault. Any science of seeing based upon such an unstable foundation must, therefore, lack validity. Since these data have been the justification for specifying what appear to be excessively high levels of illumination intensity, we must reject such specifications unless justified by valid evidence from new experimentation.

LIGHTING CODES

School lighting. The American Recommended Practice of School Lighting (35) specifies the following minimum foot candles in service: 15 for classrooms, shops and offices; 25 for sewing and drafting rooms; and 30 for sight-saving classes. There is general agreement on the importance of hygienic illumination in reading and study situations. The recommended foot candle levels seem satisfactory in view of research findings other than those cited in the code. There should be, of course, a sound experimental basis for recommendations of this kind. Tinker (23) has pointed out that the recommended practice for school lighting is based upon conclusions derived from misinterpreted experimental results. Fortunately, the recommended practice is adequate in spite of inferences from inadequate data.

In a later publication by Sturrock (21), the foot candle levels are not in an approved code but are listed as the levels found desirable in the experience of successful business institutions, i.e., good present-day practice. For schools the foot candles listed include: 30 for study halls, class rooms, general laboratories, general manual training; 50 for drawing room, close work in laboratory, sight saving classes; 100 (considered especially low) for close work in manual training, and in sewing rooms. It is obvious to the impartial person who knows the field that these suggestions represent more intense illumination than is necessary for adequate seeing in the school situation. Data summarized by Tinker (24) and additional experimental evidence (25, 27) indicate that about 15 foot candles are adequate for ordinary schoolroom tasks and that 25 to 30 foot candles are satisfactory for the more severe tasks. Justification for the higher intensities is sought in the discussions of Luckiesh and Moss (12, 14) and Luckiesh (10). These have been evaluated above.

Office lighting. The Recommended Practice of Office Lighting (36) includes the following foot candle levels: 50 for difficult seeing tasks such as accounting, bookkeeping, and drafting; 25 for ordinary seeing tasks such as general office work, private office work, mail rooms; 10 for casual seeing tasks such as reception rooms and washrooms; 5 for simple seeing tasks such as halls and stairways. Considering the se-

verity of the tasks performed by some workers in general offices and special (as accounting) offices, the above recommendations are satisfactory. The 50 foot candles, however, should be considered liberal even for the difficult seeing tasks. The statement that "higher values will contribute greatly to accuracy, speed and ease" cannot be accepted as valid.

Sturrock's (21) summary of good present day practice does not deviate markedly from the recommended practice except that typing and prolonged reading of shorthand notes are listed at 50 foot candles and intermittent reading and writing at 30 foot candles. Each of these is about twice what is needed in terms of the visual task. The basis for the higher intensities is in terms of the discussions of Luckiesh and Moss (12, 14). The inadequacy of these data has been pointed out above.

Industrial lighting. A wide range of illumination intensities is recommended for various tasks in industry (37). Among the higher foot candle recommendations are: *over* 100 foot candles for such operations as extra fine assembly, automobile finishing and inspecting, cutting and sewing dark goods, engraving, proofreading, final inspection of tire casings, grading and sorting tobacco products, and certain inspection work in textiles; 50 to 100 foot candles for such operations as automobile assembly line, glass works inspection, fine inspection, bookkeeping, font assembly-sorting in printing industry, tin plate inspection, and stitching dark leather. With regard to all the recommendations, one is cautioned that the foot candles are minimum operating values and that in almost every instance higher values may be used with greater benefit.

It is stated that the recommendations are taken from a series of studies on the illumination needs of specific industries, or, if not available there, from current good practice. Examination of these studies (listed on page 23 of the report) indicates that in the main they are surveys rather than experiments. Furthermore, there is a lack of adequate descriptions of the survey techniques employed. In a few instances a general description of methods was given. Apparently what happened was first to make a survey of practice. This was followed by some sort of job analysis to determine what had to be discriminated. Then by reference to research studies (such as those reported by Luckiesh and Moss in their books) the intensity level of illumination presumably needed for the specific job was deduced. This method has some virtue provided sound data are referred to, which was not done in these cases. In a few instances it is stated that visibility measurements were made. Occasionally installations to achieve the recommendations

were made, the effect observed and additional modifications made. In no case was there experimental determination of the light intensity needed.

There are no valid experimental data which indicate that more than 50 foot candles are needed even for those practical visual tasks which approach threshold discrimination. Furthermore, as pointed out by Harrison (8), visual comfort may decrease under high intensities.

Home lighting. The most recent recommended practice for home lighting (38) specifies intensities ranging from 10 foot candles on card tables to 100 and more for sewing on dark goods. Forty foot candles are recommended for such situations as children's study table, kitchen work counter, laundry, and for prolonged reading. There is no valid reason for going above 25 to 30 foot candles for the more severe visual tasks in the home (24). Approximately 15 foot candles is adequate for many of these visual tasks. Figure 1 in *Recommended Practice of Home Lighting* (38) is misleading. "This chart shows the extent to which occupations and poor seeing conditions leave their mark on eyesight." The implication is that poor illumination causes ocular disability. There are no valid data which indicate this to be so. This chart represents an unjustified form of propaganda.

Present-day practice. Sturrock (21) has assembled foot candle levels of illumination which are labeled "good present-day practice." The tables are preceded by a classification (after Luckiesh and Moss) of foot candle needs for visual discrimination of tasks varying in difficulty. The material is apparently designed as a guide but is not necessarily in the form of recommendations. This sort of thing is valuable in many ways. But since it is based to a considerable degree upon the material presented by Luckiesh and Moss (12, 14) and by Luckiesh (10), the illumination intensities are excessively high in some instances—as 100 foot candles for sewing and proofreading, and 50 foot candles for reading small type and for kitchen counters. It should be pointed out, however, that much of the material is fairly satisfactory.

In general, recommended practice prior to 1940 (35) is fairly adequate, but as new codes are issued at later dates the apparent tendency has been to recommend as intense lighting as the traffic will bear. This is justified by referring to the work of engineers (largely Luckiesh and Moss) who state that these high intensities are nevertheless inadequate for easy seeing. As pointed out above, both the experiments and the conclusions which are cited as fundamental are frequently invalid. Furthermore the data are out of line with other independent experimental results.

VISUAL FACTORS

Eye disabilities. It is generally accepted that eyes with disabilities, even when corrected by glasses, need brighter light than normal eyes for adequate visual discrimination. Ferree and Rand (6) and Ferree, Rand and Lewis (7) are usually cited as supporting evidence. In the first study (6), it was found that apparent diopters of accommodation increased more for 14 presbyopes than for normal eyes in going from 1 to 5 to 25 foot candles of light. Interpolation indicates that for the normal eyes the curve of improvement shows little rise after about 8-10 foot candles; for the presbyopes, after about 15 foot candles. In addition, *one* myope and *one* presbyope were compared with a normal subject by measuring apparent diopters of accommodation at 13 intensities from 0.5 to 100 foot candles. The curve of efficiency for the normal person improved rapidly to 5 foot candles, then more slowly to about 20 and very gradually thereafter; for the myope there was considerable improvement to about 20 foot candles and little thereafter; for the presbyope there was considerable improvement to about 38 foot candles, and then slower improvement to 100 foot candles. It is of course impossible to generalize from *one case*, but apparently those with eye disabilities need somewhat brighter light than normals for clear seeing. This does not mean that they need 100 foot candles or more, as some people wish to imply.

In the other study (7) Ferree, Rand and Lewis were concerned with distant (20 feet) vision. The visual acuity for 4 presbyopes was compared with acuity for 3 normal people. The presbyopes continued to gain in visual acuity from 25 to 100 foot candles while the normal eye made little gain within this range. Since there is little or no relation between acuity of distant vision and acuity at near vision, these results have no bearing upon visual discrimination at the work surface (desk, work bench, etc.). Furthermore, one should not prescribe illumination for suprathreshold tasks in terms of threshold measurements (visual acuity). There is no evidence from these studies which implies that excessively high foot candles are necessary for those with ordinary visual disabilities. Rather, they suggest a moderate increase for those with corrected vision as compared with normal eyes.

Visual adaptation. It is well established that the eyes readily adapt to easy and effective seeing over a wide range of illumination intensities. This adaptation is rather slow in going from bright to dimmer illumination (for practical purposes, 15-20 minutes) and rapid in going from dim to bright illumination (1-3 minutes). Tinker (25) has demonstrated

that when adaptation is incomplete on shifting to a lower level of illumination, speed of perception is retarded. When adaptation is adequate, however, visual perception in reading is fully effective from 3 foot candles up for normal eyes in reading legible print. In another study, Tinker (26) showed that subjects tend to prefer for reading approximately the illumination intensity to which they have been adapted, whether it be 8 or 52 foot candles. These data indicate that readers tend to consider comfortable for easy reading any one of a wide range of illumination intensities provided such intensities are above critical levels and provided visual adaptation is adequate. Codes of lighting have consistently ignored the role of visual adaptation in seeing. They carefully point out that the eye has evolved under the bright illumination of daylight, but do not mention that the eye also evolved to see adequately at low as well as at high intensities of light.

ILLUMINATION FOR ADEQUATE SEEING

Critical levels of illumination. The critical level of illumination is the intensity beyond which there is no further increase in efficiency of performance as the foot candles become greater. Tinker (24) has summarized the data for critical levels of illumination: for reading of legible print (about 10 point on good paper) by adults, it is approximately 3 to 4 foot candles; for reading and study of children, 4 to 6 foot candles; for arithmetical computations, less than 9.6 foot candles; for sorting mail, 8 to 10 foot candles; for the exacting task of setting six-point type by hand, 20-22 foot candles; and for very fine discrimination required to thread a needle, 30 foot candles. In a later study, Tinker (27) found the critical level of illumination for reading newspaper print to be about 7 foot candles. Employing intensities from 2 to 55 foot candles, Rose and Rostas (20) found that reading efficiency, in terms of speed and comprehension, did not increase by a measurable amount with increased intensity of illumination.

Adequate levels of illumination. It is obvious that visual work should not be done at critical levels of illumination. There should be an adequate margin of safety to provide for individual variation and the like. For such visual tasks as reading good-sized print (10 to 11 point) on a good quality paper, i.e., print of good legibility, 10 to 15 foot candles should provide hygienic conditions when one's eyes are normal. For situations comparable to the reading of newsprint, 15 to 20 foot candles should be adequate. In situations involving the reading of handwriting and other comparable tasks, 20 to 30 foot candles seem desirable. For

tasks comparable to discrimination of 6 point type, there should be 30 to 40 foot candles. And for the most severe tasks encountered in work-day situations, 40 to 50 foot candles will be found adequate. There is no valid experimental evidence now available that indicates a need for over 50 foot candles intensity for adequate visual discrimination. The intensity values from 10 to 20 should be increased somewhat (5 to 10 foot candles) for eyes with slight disabilities or for those with corrections. For the higher values, however, no practical gain will be achieved for these people by increasing the intensity. The above suggestions hold for school children as well as for adults. In general, the child has much less severe visual tasks than adults.

Intensity of illumination cannot be prescribed without coordinating it with other factors such as distribution of light and brightness contrast. A good example of the uselessness of excessively bright light is found in the study by Darley and Ickis (4). They were concerned with vision in the drafting room, a very severe visual task. In comparing 30 with 75 foot candles of indirect light, they found the efficiency ratings for the two to be only slightly different. When they compared 40 with 80 foot candles of direct light (troffer) under conditions of no reflected glare, they also found no significant differences in the efficiency ratings. The observations of Harrison (8) are relevant here. He points out the danger of glare with installations of 50 foot candles and above of artificial illumination.

SUMMARY

Examination of the literature upon which lighting recommendations are based reveals that some techniques of experimentation are invalid, and that interpretations from certain other data are unwarranted. Some of the recommendations are adequate, others are not. The trend seems to be to specify as high intensities as the traffic will bear and at the same time to suggest to the consumer that if he uses still higher intensities he will improve his ease of seeing. All will agree that there should be sufficient light for adequate seeing. It is high time, however, that the consumer know what is adequate and what is surplus. As pointed out by Winslow (34), illumination should conform to real human needs. It is human health and comfort which are at stake.

In general the recommended practice concerning distribution of light, brightness contrast and color of light is satisfactory.

BIBLIOGRAPHY

1. BITTERMAN, M. E. Heart rate and frequency of blinking as indices of visual efficiency. *J. exp. Psychol.*, 1945, 35, 279-292.
2. BITTERMAN, M. E., & SOLOWAY, E. The relation between frequency of blinking and effort expended in mental work. *J. exp. Psychol.*, 1946, 36, 134-136.
3. BITTERMAN, M. E., & SOLOWAY, E. Frequency of blinking as a measure of visual efficiency: some methodological considerations. *Amer. J. Psychol.*, 1946, 59, 676-681.
4. DARLEY, W. G., & ICKIS, L. S. Lighting and seeing in the drafting room, *Illuminating Engineering*, 1941, 36, 1462-1487.
5. EAMES, T. H. Review of M. Luckiesh and F. K. Moss, *Reading as a visual task*, *Columbia Optometrist*, 1942, 16, 7.
6. FERREE, C. E., & RAND, GERTRUDE. The effect of intensity of illumination on the near point of vision and a comparison of the effect for presbyopic and non-presbyopic eyes. *Transactions I.E.S.*, 1933, 38, 590-611.
7. FERREE, C. E., RAND, GERTRUDE, & LEWIS, E. F. The effect of increase of intensity of light on the visual acuity of presbyopic and non-presbyopic eyes. *Transactions I.E.S.*, 1934, 29, 296-313.
8. HARRISON, W. What is wrong with our 50 foot-candle installations? *Transactions I.E.S.*, 1937, 32, 208-223.
9. HOFFMAN, A. C. Luckiesh and Moss on reading illumination. *J. appl. Psychol.*, 1947, 31, 44-53.
10. LUCKIESH, M. *Light, vision and seeing*. New York: Van Nostrand, 1944.
11. LUCKIESH, M., & MOSS, F. K. *The new science of seeing*. Cleveland: General Electric Co., 1934.
12. LUCKIESH, M., & MOSS, F. K. *The science of seeing*. New York: Van Nostrand, 1937.
13. LUCKIESH, M., & MOSS, F. K. Illumination and eye health. *New Eng. J. Med.*, 1941, 224, 1117-1119.
14. LUCKIESH, M., & MOSS, F. K. *Reading as a visual task*. New York: Van Nostrand, 1942.
15. LYTGOE, R. J. *The measurement of visual acuity*. London: His Majesty's Stationery Office, 1932.
16. MACPHERSON, S. J. The effectiveness of lighting—its numerical assessment by methods based on blinking rate. *Transactions I.E.S.*, 1943, 38, 520-522.
17. MCFARLAND, R. A., KNEHR, C. A., & BERENS, C. Metabolism and pulse rate as related to reading under high and low levels of illumination. *J. exp. Psychol.*, 1939, 25, 65-75.
18. MCFARLAND, R. A., HOLWAY, A. H., & HURVICH, L. M. *Studies of visual fatigue*. Boston: Graduate School of Business Administration, Harvard Univ., 1942.
19. McNALLY, H. J. The readability of certain type sizes and forms in sight-saving classes. Teachers College, Columbia Univ., Contrib. to Educ., No. 883. New York: Bureau of Publications, Teachers College, Columbia Univ., 1943.
20. ROSE, F. C., & ROSTAS, S. M. The effect of illumination on reading rate and comprehension of college students. *J. educ. Psychol.*, 1946, 37, 279-292.
21. STURROCK, W. Levels of illumination. *Magazine of Light*, 1945, 14, No. 4, 26-36.
22. TINKER, M. A. Cautions concerning illumination intensities for reading. *Amer. J. Optom.*, 1935, 12, 43-51.
23. TINKER, M. A. The new "standards" for school lighting. *Sch. & Soc.*, 1939, 49, 95-96.

24. TINKER, M. A. Illumination standards for effective and comfortable vision. *J. consult. Psychol.*, 1939, 3, 11-20.
25. TINKER, M. A. The effect of illumination intensities upon speed of perception and upon fatigue in reading. *J. educ. Psychol.*, 1939, 30, 561-571.
26. TINKER, M. A. Effect of visual adaptation upon intensity of light preferred for reading. *Amer. J. Psychol.*, 1941, 54, 559-563.
27. TINKER, M. A. Illumination intensities for reading newspaper type. *J. educ. Psychol.*, 1943, 34, 247-250.
28. TINKER, M. A. A reply to Dr. Luckiesh. *J. appl. Psychol.*, 1943, 27, 469-472.
29. TINKER, M. A. Review of M. Luckiesh and F. K. Moss, *Reading as a visual task*. *J. appl. Psychol.*, 1943, 27, 116-118.
30. TINKER, M. A. Illumination intensities preferred for reading with direct lighting. *Amer. J. Optom. & Arch. Amer. Acad. Optom.*, 1944, 21, 213-219.
31. TINKER, M. A. Review of M. Luckiesh, *Light, vision and seeing*. *J. appl. Psychol.*, 1945, 29, 252-253.
32. TINKER, M. A. Validity of frequency of blinking as a criterion of readability. *J. exp. Psychol.*, 1946, 36, 453-460.
33. TINKER, M. A. Illumination standards. *Amer. J. Pub. Health*, 1946, 36, 963-973.
34. WINSLOW, C. E. A. How many foot candles? *Amer. J. Pub. Health*, 1946, 36, 1040-1041; also reprinted in *J. appl. Psychol.*, 1947, 31, 140-142.
35. *American recommended practice of school lighting*. New York: Illuminating Engineering Soc. & Amer. Inst. Architects, 1938. Pp. 60.
36. *Recommended practice of office lighting*. New York: Illuminating Engineering Soc., 1942. Pp. 47.
37. *American recommended practice of industrial lighting*. New York: Illuminating Engineering Soc., 1942. Pp. 51.
38. *Recommended practice of home lighting*. New York: Illuminating Engineering Soc., 1945. Pp. 40.

ON FESTINGER'S EVALUATION OF SCALE ANALYSIS

LOUIS GUTTMAN

Department of Sociology and Anthropology, Cornell University

The theory of scale analysis had its origin some seven years ago. Since that time, especially by virtue of extensive and intensive research done in the Army, some of its further ramifications have been explored and several techniques have been devised for carrying an analysis out in practice. The power and incisiveness of this approach have been demonstrated in numerous attitude and opinion surveys made in the past several years, as well as in studies of achievement tests. A pleasing feature has been the simplicity of the techniques involved.

Most of the material, with respect to both applications and theoretical developments, is as yet unpublished. A manuscript has been prepared by Edward A. Suchman and the writer which will give the first comprehensive statement of both the theory and practice of scale analysis. This manuscript will form part of the four volumes soon to be published by the Social Science Research Council on the work of the Research Branch, Information and Education Division of the War Department. These volumes will also provide many illustrations of how scale analysis has been used for practical problems. Meanwhile, some brief statements of the principal concepts and instructions for practical procedures are available in article form to those who wish to use this approach in their own research (see the bibliography below).

On the basis of some articles which have been published and of some mimeographed progress reports, Festinger (1) has recently attempted a survey and evaluation of scale analysis. Since his survey is not based on all the information available, it is admittedly tentative and incomplete. In addition, full advantage has not been taken of the material which Festinger used as his sources; he raises a number of points which have already been answered there, and also introduces erroneous interpretations and conclusions.

It seems worthwhile to discuss at the present time some of Festinger's criticisms in order to help clarify the issues and to correct some important misapprehensions. Attention is also called to some articles that have appeared since Festinger prepared his paper, discussing various aspects of scale analysis (10, 11, 13).

Three of Festinger's points will be analyzed here: (a) criteria for scalability, (b) techniques of analysis, and (c) the use of scale analysis in practice. In the course of the discussion, some other aspects will be brought out which Festinger has not considered.

CRITERIA FOR SCALABILITY

Reproducibility. The main purpose of scale analysis is to test the hypothesis that a universe of qualitative items can be represented by a quantitative variable. In order for the universe to be represented exactly by a quantitative variable, each item must be a perfect function of that variable, or be perfectly reproducible from it. Thus the concept of reproducibility is paramount in scale analysis.

In practice, only a sample of items is used from the universe of content. Furthermore, in practice, it is not expected to find perfectly reproducible or scalable universes. Among other things, perfect reproducibility implies perfect test-retest reliability, which is certainly not to be expected empirically. However, if the reproducibility of the entire universe is very high, say over 90%, then that may be sufficient for many practical purposes. A quantitative variable which will represent an indefinitely large universe of items that well will ordinarily not lose much predictive power, whether used for predicting outside variables or whether predicted from outside variables. This will especially be true if the errors of reproducibility are random.

Since universe reproducibility must be estimated on the basis of only a sample of items, it becomes evident that the sample's reproducibility alone may not be a sufficient guide. Festinger criticizes the sample reproducibility coefficient for its inadequacy.* This inadequacy was recognized at the outset in scale analysis. The same kind of examples that Festinger uses (1, pp. 156-157), showing how five or nine statistically independent items can have high reproducibility, were worked out previously; several such examples will appear in the forthcoming volume. Indeed, there is an even worse case than that of statistical independence, namely that wherein some items have *negative* relationships with others; this is worse than being statistically independent from the point of view of scale analysis. Examples can be constructed showing how even in this case it is possible to have suprisingly high reproducibility in a small sample of items.

Festinger omits to point out that this problem about reproducibility was raised before, and that several answers have already come forth. In one of my mimeographed reports to which Festinger refers (15), there is the following question and answer:

Q. *Is reproducibility by itself a sufficient test of scalability?*

A. No. It is the principal test, but there are at least three other features

* Hausknecht (12) has raised this criticism earlier, also without taking cognizance of the fact that other criteria have always been used as discussed below.

that should be taken into account: (a) range of marginals, (b) random scatter of errors, (c) number of items in the sample.

Further questions and answers elaborate on the point. And again, in another paper (7) to which Festinger refers, it is stated:

The percent reproducibility alone is not sufficient to lead to the conclusion that the universe of content is scalable. The frequency of responses to each separate item must also be taken into account for a very simple reason. Reproducibility can be artificially high simply because one category in each item has a very high frequency. It can be proved that the reproducibility of an item can never be less than the largest frequency of its category, regardless of whether the area is scalable or not.

And further:

An empirical rule for judging the spuriousness of scale reproducibility has been adopted to be the following: No category should have more error in it than non-error.

If this latter rule alone were applied to Festinger's examples, it would immediately reject the hypothesis that the items are from scalable universes. The consideration about pattern of error would also disqualify the hypothesis that the items were from scalable universes.

An Alternative. One contribution to spuriously high reproducibility is the fact that each item is being related to a score which is based in part on the item. An alternative way to compute the coefficient of reproducibility is to hold out each item in turn from the analysis, thus obtaining N sets of trial scale scores. The errors for each item can then be counted from its relationship to the score based on the $N-1$ other items.

If this *partial-score* method were used on *statistically independent* items, then the reproducibility for each item would be precisely the relative frequency of its modal category. Thus, in Festinger's example (1, p. 156) of five independent dichotomies with marginals 80%, 60%, 50%, 40%, and 20% the respective modal relative frequencies are 80%, 60%, 50%, 60%, and 80%; hence, the reproducibility of all five items, computed from partial scores, would be the mean of the latter five percentages or 66%, compared with the spurious 86% Festinger obtained from whole scores. Indeed—no matter what the interrelations of the five items were—their reproducibility could not be less than 66%, because reproducibility of an item can never be less than its modal frequency. Similarly, in Festinger's second example (1, p. 157) of nine statistically independent dichotomies with marginals .9, .8, .7, .6, .5, .4, .3, .2, .1, the respective modal proportions of the items are .9, .8, .7, .6, .5, .6, .7, .8,

and .9, so the reproducibility of the set cannot be less than .72; Festinger finds .83 reproducibility from whole scores whereas if part scores were used the obtained reproducibility would be .72.

Items with extreme marginals like .9 and .1 do not help much in testing reproducibility since such items can never have more than 10% error.

In practice, it does not usually seem worthwhile to bother with partial scores, although this technique is available for doubtful cases. The fictitious examples of independent items do not illustrate what is to be expected in practice. Attitude (or achievement) items of the same general content are usually sufficiently correlated so that scores based on eleven of them will not be substantially different from scores based on twelve. Reproducibility from whole scores will not be much greater than from part scores—so their spurious excess of reproducibility over that from part scores can be largely ignored. Furthermore, even part-score reproducibility is not a sufficient test of scalability, for the additional criteria mentioned above must also be considered.

There is room for more improvement on criteria for scalability when samples of content are used, but it should be made clear that reproducibility by itself has not and is not the sole basis for drawing inferences from a sample of items. It is the basic one, because the reproducibility of the universe is essentially what is in question, but additional criteria have been and are being used.

Reliability. The suggestion that Festinger makes that the expected occurrence of scale responses be calculated under the assumption of a perfect scale plus a certain degree of unreliability is a promising one. This idea had been thought of in the earlier stages of the development of scale analysis but discarded in the form Festinger has suggested. The proportion of people with no scale errors *cannot* be properly calculated by the method that Festinger uses. Apparently he assumes that if .9 is the proportion of population responses that will be in the scale pattern for one question, then the proportion that will be jointly *within* the scale pattern for seven questions is $(.9)^7$ or 47.8%. Unfortunately, the same reasoning would say that the proportion of people who have seven responses *outside* the scale pattern should be $(.1)^7$; and in general the proportion of people with X scale responses and $7 - X$ scale errors should be given by the binomial distribution

$$\frac{7!}{X!(7 - X)!} (.9)^X (.1)^{7-X}.$$

But this is impossible, for nobody can have *all* his responses as scale

errors. Indeed, for the empirical example that Festinger borrowed (1, Fig. 2, p. 157), no matter what pattern of response a person may have, he can be placed into one of the scale patterns with at most four errors. Therefore, the range of possible errors for each person is 0 through 4, rather than 0 through 7 as Festinger supposes. This means that Festinger's calculations cannot be carried out consistently to estimate reproducibility under the given assumption. The difficulty is that whether a person will fall into the scale pattern is *not* independent of whether another of his responses is within the scale pattern. Unreliability does not behave that way with respect to the scale pattern.

The actual reproducibility of this example of seven questions was about .85 rather than the .9 Festinger assumed. It is interesting to note that $(.85)^7$ is .32, which is not far from the "over one-fourth" perfect scale types reported. Actually, the universe sampled by these seven questions would not now be accepted as sufficiently scalable but would be broken up into sub-universes; the study was made when 85% reproducibility was the empirical rule rather than the present 90%. The study did serve its purpose well, however, as collateral evidence presented there showed.

The further calculation that Festinger makes of adding 3.7% to his 47.8% seems based on an unfortunate double usage of the word "chance." In his second paragraph on p. 158 (1), "chance" is used to mean statistical independence between items. Such independence cannot exist simultaneously with the assumption of a scale pattern in his following paragraph; that is, the 7% who fall into perfect types under the hypothesis of independence of items have nothing to do with the distribution of error under the assumption of uni-dimensionality plus unreliability. The binomial distribution by itself—if it were correct—takes care of the second situation. Hence, Festinger's calculations are incompatible in adding 3.7% (7% of $1 - .478$) to 47.8% to obtain 52.2% as the "chance" proportion. The 7% is correct for independent items; the 47.8% would be correct for the scale-plus unreliability case if the binomial hypothesis held; and the two cases do not hold simultaneously. "Chance" means something different in each case.

A consistent use of reliability. Several correct approaches to the use of the concept of unreliability are possible, instead of the inconsistent binomial approach. One such approach will be sketched here briefly for the case of dichotomies. Let n be the number of dichotomies in the sample of items so that there are $n+1$ scale types or ranks possible. Let r be the rank of the type that is "positive" on r of the items; r ranges from 0 to n . Let Pr be the proportion of the population whose

"true" rank on the n items is r , and let Prj be the probability a person of "true" rank r will be "positive" in the j th item ($j=1,2,\dots,n$). There are 2^n types of people—scale and non-scale—possible on the n dichotomies. The expected proportion in each of the 2^n types can be calculated from the Pr and Prj . Conversely, from the observed 2^n proportions in an actual experiment, the Pr and Prj can be estimated. There are $n+1$ parameters Pr , of which n are independent since their sum must be unity. There are $(n+1)n$ parameters Prj , all of which are independent. Hence, there are $n+(n+1)n$, or $n(n+2)$, independent parameters to be estimated from 2^n-1 independent observations. If n is greater than 5, this provides more equations than there are unknowns—so the hypothesis of the scale structure can be tested, as well as having the parameters estimated. Unfortunately, the equations involved in the above analysis are curvilinear, and do not seem to lend themselves to practical use because of the difficulties in the numerical computations. Furthermore, even this analysis has been simplified by assuming that persons within the same "true" rank were equally reliable within each item. Without this simplifying assumption, the equations would have innumerable more parameters.

In any analysis using the concept of test-retest reliability, it must be remembered that scalable data must in general be highly reliable, although the converse is not necessarily true. The coefficient of reproducibility—especially if computed by the part-score technique described above—sets a *lower bound* to the average reliabilities of the items (6, and especially 8). In particular, if items are perfectly reproducible, they are perfectly reliable. Hence, Festinger errs in his assertion that "Even if a perfect scale were achieved these claims [concerning invariance properties] would all be limited by the degree of reliability . . . of the questions asked" (1, p. 160). Perfectly scalable data are perforce perfectly reliable. Conversely, highly unreliable data cannot be scalable. One of the contributions of a scale analysis is to provide automatically information about reliability by helping set a lower bound to it for each item.

The simple criteria used in conjunction with that of reproducibility for sample data do serve to distinguish between data that are highly scalable and those that are not. The case where the items are independent will always be rejected on the basis merely of the criterion of improvement, namely, that no category should have more errors than non-errors. The further criteria of studying patterns of error also tend to insure that no dominant second variable is present even if reproducibility is high. That is what is meant by the statement that "in imperfect

scales, scale analysis picks out deviants or non-scale types for case studies." If no non-scale types have substantial frequencies, then that tends to indicate that there is no substantial second factor present. However, if one or more non-scale types do have a substantial frequency, then that is an indication of where an additional factor (or factors) is entering into the picture. If an additional factor is sufficiently prominent, it may be worthwhile to try to piece it out further by asking additional questions. The universe might be divided into two or more sub-universes, each of which may be scalable separately. Or it may turn out that the additional factor is so highly correlated with the most dominant factor that it does not make much difference whether they are treated as two separate variables or as a single variable.

The problem is not to find out whether a perfect scale is present in practice, but rather whether it is worth worrying about any additional variables that may be present. The criteria used in practice are believed to provide an answer to this and to decide properly whether or not a set of data can be regarded as sufficiently scalable for most practical purposes.

Quasi-scales. One kind of non-scalable universe is called a *quasi-scale*. A quasi-scale is different from a scale, not just in the reproducibility, but in *the entire pattern of responses*. Festinger seems to have misunderstood the definition of a quasi-scale, for he seems to believe that it differs from a scale only with respect to reproducibility (1, p. 156 and p. 159). A universe which is quasi-scalable will ordinarily have less than 85% or so reproducibility, but that is not its distinguishing feature. The distinguishing feature is *the gradient in the responses to the items*. Cutting points cannot be established (as in the case of a scale) which will enable one to say that a person above the point is in one category of an item and a person below the point is in another category; but one can state that, if one person is higher than another in the quasi-scale, then his probability of being in a higher category of an item is correspondingly greater.

There are many kinds of configuration which are less than 85% or 90% reproducible and which are not quasi-scales at all. For example, an area may have two or more dominant factors in it, in which case it would not be either a scale or a quasi-scale. In a quasi-scale, there are one dominant factor and infinitely many small factors. The order of people in a quasi-scale is according to the dominant factor, and is essentially invariant from sample of items to sample of items, provided that the samples are large enough. There is a great deal of work yet to be done on the theory of quasi-scales, but enough is known to say that

they have quite a different character from scales and from other kinds of universes. Another distinguishing feature between a scale and a quasi-scale is that the scale has an intensity function and further meaningful components, whereas a quasi-scale does not have an intensity function or further components of that kind.

Neurotic phenomena have been found to be quasi-scalable. For example, the Neuro-Psychiatric Screening Adjunct, which is the official paper and pencil test used at all military stations since October, 1944, is a quasi-scale and is a product of a rigorous investigation of efficient screening tests made possible by the scale analysis approach (16, 17).

TECHNIQUES FOR SCALE ANALYSIS

Scalogram devices. There are several alternative schemes now available by which to carry out a scale analysis in practice. They are virtually equivalent in terms of the results they yield, but they differ somewhat in operation. *Scalogram boards* have been the principal device used by the War Department, and are perhaps the most flexible and easiest to use. The boards are relatively simple to make and to operate; the cost depends upon how large a board is desired and whether or not a pair is to be made. If a single board is used instead of two, then the workmanship need not be precise and the board can be made fairly cheaply by any carpenter. There are alternative mechanical schemes that might be used instead of the wooden board, and undoubtedly other schemes will be invented in the future which will be even easier to construct. Instructions for the construction and use of a scalogram board will appear in the forthcoming volumes on the work of the Research Branch.

The Cornell technique (7) is also very easy to learn; it is taught in a course on attitude and public opinion analysis to students who have no background whatsoever in statistics. For achievement tests, where all items are dichotomous—being marked either right or wrong—the Cornell technique is perhaps the best of all to be used. For dichotomies, there is no problem of combination of categories, so that there is but one trial to be made in an analysis. The Cornell technique suffers a bit in flexibility compared to the scalogram board when a series of trials has to be made. Ordinarily, but two trials may be needed in an analysis, and the Cornell technique has proved very advantageous in such cases for general research purposes. It can be carried out on IBM equipment as well as by hand.

The Goodenough technique (2) is based upon an explicit tabulation of all combinations of responses that actually occur. It is more "rigor-

ous" than the preceding two techniques in that it counts the errors at each stage. However, it yields no different results in the end. Apparently Festinger has not worked through the Goodenough technique to see how it does work out in practice.* The first step seems simple, but it takes a good deal of experience to master the three following steps. The process becomes very bulky and involved when ten or twelve items are used.

The Cornell Technique has the advantage that its complexity does not at all change, regardless of the number of items (though of course the amount of labor increases with the number of items). The same lack of increase of complexity holds to a slightly less degree with the scalogram board.

The problem of metric. The earliest technique for scale analysis was that of least squares (3). It is quite properly to be abandoned as a procedure in practice because it is certainly far more cumbersome than the others. However, the equations involved have turned out to be of basic importance in *interpreting* a scale, and have led in particular to the empirical treatment of the intensity function which is proving so vital for attitude and public opinion work. Also, the basic thinking behind the equations have led to a solution to the related problem of paired-comparisons (12).

In the beginning of my work on scale analysis, I had thought that one of the most important problems was that of metric. I had thought that how to obtain weights for items was perhaps the leading problem to be solved. But as the theory of scale analysis developed, it became clear that the problem of weights was essentially a minor one for most practical purposes. Indeed, for the *perfect* scale pattern, it is easy to see that if scores are to be obtained for people by adding up weights assigned to categories of items, then, no matter what weights are used—as long as they have the proper rank order within each item—the scores of the people will have *exactly* the same rank order. The ordering of people in this sense does not depend at all upon finding a particular weighting system.

The important problem turned out to be that of finding the *structure* a universe of items must have in order to be scalable; it was not that of finding weights.

* Festinger also apparently has misread Goodenough as to how to measure reproducibility. Goodenough explicitly says that "at least 85% of the total number of responses must fall within the scale pattern, so that it is possible to reproduce 85% correctly all the responses of all the respondents from the scale scores" (2, p. 184). Festinger seems to have misread this to mean that 85% of the individuals fall into perfect scale types.

The problem of a metric does turn out to come into the picture for further problems, and it first appeared as a practical problem with respect to that of bias in questionnaire wording (11, 13, 14). The problem here was, after people are ranked from a high to a low on an attitude or opinion, to find a dividing point in the order such that the people on one side can be called positive and people on the other side can be called negative. The equations of scale analysis, when applied to the perfect scale pattern, show a most remarkable result. They show that a universe of items which is perfectly scalable can be resolved into an infinite series of principal components, the first of which provides the basic metric, the second of which is the intensity component, and the remaining ones are as yet not named (10). Empirical study of the intensity function has afforded for the first time a scientific solution to the problem of question bias.

These equations, then, show that a scalable attitude is somewhat different from the twelve-inch ruler that Festinger uses as an analogy (1, p. 160). The responses of a person to items in a scalable universe are seen by means of these equations to be a function of the person's metric score, his intensity, and the further components in the scale. The person's rank order is sufficient to reproduce his responses exactly; in this sense, the responses of the population are but a function of a single variable. Resolving the responses into components by the alternative device of the least squares equations shows the responses to be a function of infinitely many variables, each of which is a function of the rank order.

These striking results from using the least squares equations in conjunction with the perfect scale pattern will be elaborated on in the forthcoming publication on the work of the Research Branch. It might further be pointed out here that these equations resolve also the paradox which appears in achievement tests where the *difficulty* of an item seems to introduce a factor different from the common *content* factor that the items may have. Since scale analysis applies to achievement tests as well as to attitude or opinion areas, achievement tests also are resolvable into the principal components of a scale. In a scalable achievement test, then, each item is a function of but a single dimension from the point of view of reproducibility, but a function of infinitely many dimensions from the point of view of principal components. The apparent contradiction between these two points of view is resolved by the fact that the infinitely many principal components in turn are perfect functions of the rank order of people.

USES OF SCALE ANALYSIS

Incidence of scales. The theory and techniques of scale analysis provide a test of the hypothesis that a universe of qualitative items can be represented by a single quantitative variable. This hypothesis is appropriate for any qualitative universe obtained by any method of observation. The universe may be a set of items recorded on a questionnaire, or observations obtained in non-directive interviews, by participant observation, or by any other technique of gathering data. No matter how the data are gathered, each observation is but a sample of all similar observations that could have been obtained, and the entire universe of observations is ordinarily of interest.

As Festinger suggests, scalable universes may be the exception rather than the rule. Festinger does not give any explicit reasons for his belief, but this position will be substantiated in the forthcoming volume. It has already been pointed out that one possible reason for the existence of an attitude scale is that of a homogeneous culture (4, p. 149). If a population is not subjected to the same social stimuli with respect to the attitude, it might be expected that it will prove to be unscalable for them. The fact that neurotic phenomena have not been found scalable can perhaps be explained in this fashion. Similarly, an area of achievement may be expected not to be scalable if there is no uniform program of training for the population involved.

Another reason for expecting many universes not to be scalable in practice is that the notion of a universe is so comprehensive. Each sub-universe of a universe is of course itself a universe. Since there is ordinarily a vast number of imperfectly related sub-universes, there must be a vast number of combinations of them which are non-scalable universes. Merely this formal consideration would lead one to believe that most universes are not scalable. Non-scalable universes may of course be broken down in some cases into scalable sub-universes. One of the contributions of scale analysis is to point out the need for being clear about the universe's content. By focusing on more and more homogeneous content, research can be made more meaningful and external predictions be made more effective in the long run.

The development of the above-mentioned screening test for psychoneurotics (16, 17) is but one example of how research utilizing scale analysis was more effective than it would have been had the more traditional but less incisive procedures been followed. Instead of throwing together all kinds of conceivable predictive items into one composite, fifteen different universes of content were defined which might be re-

lated to the criterion of psychoneuroticism. The structure of each of these universes was first analyzed separately. Because each was found to be either a scale or a quasi-scale, only a relatively few items from each were needed in order fully to utilize the predictive power of the universes. The multiple correlation of the criterion was then worked out on all fifteen predictors with the finding that one of the universes predicted as well as the best combination of the fifteen. This enabled the short but efficient screening test to be used with the knowledge that it retained the predictive power of innumerable many items in fifteen different universes. Such a complete usage of predictive power could not have been made without scale analysis.

From the practical point of view, another important feature here is the amount of labor saved by scalogram techniques in obtaining this maximum predictive power, compared to using more traditional techniques which are far more laborious and which would yield less effective predictions.

The two problems, that of scalability and that of external prediction are distinct but related. By focusing on the scaling problem in its own right, more effective external predictions are thereby made possible.

There are many areas which have been found to be scalable thus far, and therefore these areas can be handled economically by means of simple scale scores. Many areas have also been found not to be scalable; all such areas cannot be handled so simply. It is known how to treat quasi-scalable areas, and Lazarsfeld is now completing a theory of the latent dichotomy which also can be handled by means of a single quantification. How to utilize other kinds of non-scalable areas is still an unsolved problem. The emphasis that scale analysis makes in this connection is that unless the structure of the universe is known, it is not known how best to treat the universe for any particular purpose.

Distinction between theory and techniques. The basic *theory* of scale analysis is not to be confused with particular *techniques* for carrying out such an analysis in various kinds of situations. Festinger borders on confusing the two when he states that " 'scale analysis' seems to be an excellent technique for use with paper and pencil tests or other instances of measurement where the situation permits the inclusion of several questions centering about the same topic" (1, p. 160). If a research problem is concerned with a universe of content, then that universe must be studied. That is what the theory calls for. How adequate is the technique which Festinger implicitly advocates of studying only a single item from the universe?

One of the important aspects of a universe of content is its structure;

for example, is the universe scalable or does it have some other kind of structure? The theory of scale analysis tells what a scalable structure is, and the various properties possessed by such a structure.

The practical problem is to obtain information about the structure from only a sample of items. It has already been indicated how an adequate sample of items can be chosen to test the hypothesis of scalability. Furthermore, the number of items to be used in a pretest must be distinguished from the number of items to be used in a final study. One of the properties of a scalable universe is that only one or two items can be used in a final study for many purposes once their place in the universe is ascertained. The scalability of the universe must first be analyzed, however, by a dozen or so items in a pretest.

The statement that "most of those engaged in this type of research [public opinion] will probably find the inclusion of a series of questions which could be subjected to scale analysis not feasible from practical considerations" (1, p. 159) does not accord with what is the actual practice both in public opinion and in market research, as well as in general attitude research. It is because workers in these fields are concerned with a universe of content that they pretest various questions on the same topic; it is a foolhardy pollster who bases conclusions on but a single question. The use of the split-ballot is evidence of this concern with sampling of content. In addition, ordinary polls often include several questions on the same topic on the same ballot. The extreme position taken by advocates of "open-ended interviewing" is to ask a whole series of questions of every respondent. And of course, conventional attitude surveys almost invariably use a substantial set of questions for a given topic.

It is a misapprehension to believe that asking several questions on the same topic necessarily creates a problem of rapport. In one survey made of a national cross-section by a leading public opinion polling agency, an area of content was defined and then sampled by four questions. Some of the interviewers complained because of the great similarity of wording of questions. The questions were very similarly worded because the content concerned the size of the Navy and was very hard to discuss in different ways. But even under these adverse circumstances, the analysis was successful in showing that the area was scalable and that the zero point could be located properly by the intensity function. Even more questions in the same area had been used in the pretest in Ithaca on a cross-section of the population there, and interestingly enough there was no complaint either from the respondents or from the interviewers, although the interviewers were no different from those

used in the national cross-section and had no knowledge whatsoever of what was involved in scale analysis. An area of apparently very similar questions is an exception rather than the rule. The example about desire for post-war schooling that Festinger has borrowed (1, p. 157) certainly provides no problem of rapport, and the general run of areas studied by public opinion polls do not present any particular problem of rapport. Another large market research agency has tried scale analysis in a routine study and has found no difficulty whatsoever with it. Because of its simplicity and its objective solution to the problem of bias, this agency plans to use this approach regularly.

It seems premature, then, to conclude that scale analysis cannot be carried out in practice in public opinion work. To the contrary, scale analysis is becoming more essential in this field because it affords for the first time a scientific solution to the basic problem of bias in public opinion polls. This problem arises from the fact that a universe of content is being studied and any single question is but a sample of all possible questions that could have been asked. How can one determine which question does coincide with the zero point of the entire universe, that is, the point which divides those who are negative on the issue from those who are positive?

The intensity function provides a scientific solution to this problem (13). It provides both a definition and a technique for ascertaining a zero point for the population. Unless some such objective approach to the question of bias is used in public opinion polls, it cannot be certain how much credence to place on their reports.

By providing a solution to the problem of bias, scale analysis clears the way for asking questions in the manner which will best help establish rapport with the respondent. The particular form of a question does not affect the results of scale analysis, so the research worker can concentrate on obtaining the wording which will make the interviewing work go most smoothly. Thus scale analysis has a contribution to make toward increasing rapport in surveys rather than the contrary. Apprehension that the opposite is true seems to be due to a misconception that scale analysis presupposes a particular way of asking questions.

If progress is to be made in the scientific study of attitudes, public opinion, and achievement, it seems necessary to concentrate on the problem of the structure of content. Techniques are not worth much if not guided by any theory. The theory of scale analysis happens to lend to simple and practical techniques. To compare these techniques with others, one would have to ask: what theory of structure guides the alternative techniques and how adequately is this theory served thereby?

BIBLIOGRAPHY

1. FESTINGER, L. The treatment of qualitative data by "scale analysis." *Psychol. Bull.*, 1947, 44, 149-161.
2. GOODENOUGH, W. H. A technique for scale analysis. *Educ. psychol. Msmt.*, 1944, 4, 179-190.
3. GUTTMAN, L. The quantification of a class of attributes: A theory and method of scale construction. In P. Horst *et al.*, *The prediction of personal adjustment*, Soc. Sci. Res. Council Bull. No. 48, 1941, 319-348.
4. GUTTMAN, L. A basis for scaling qualitative data. *Amer. sociol. Rev.*, 1944, 9, 139-150.
5. GUTTMAN, L. Scale and intensity analysis for attitude, opinion, and achievement. (Mimeographed, 1945. To appear in the proceedings of the conference on military contributions to methodology in applied psychology to be published by the Univ. of Maryland Press.)
6. GUTTMAN, L. A basis for analyzing test-retest reliability. *Psychometrika*, 1945, 10, 255-282.
7. GUTTMAN, L. The Cornell technique for scale and intensity analysis. (Mimeographed, 1946. To appear in the proceedings of the conference on measurement of consumer interest to be published by the Univ. of Pennsylvania Press; also in *Educ. psychol. Msmt.*)
8. GUTTMAN, L. The test-retest reliability of qualitative data. *Psychometrika*, 1946, 11, 81-95.
9. GUTTMAN, L. An approach for quantifying paired comparisons and rank order. *Ann. math. Statist.*, 1946, 17, 144-163.
10. GUTTMAN, L. Suggestions for further research in scale and intensity analysis of attitudes and opinions. *Int. J. Opin. & Att. Res.*, 1947, 1, 30-35.
11. GUTTMAN, L., & SUCHMAN, E. A. Intensity and a zero point for attitude analysis. *Amer. sociol. Rev.*, 1947, 12, 57-67.
12. HAUSKNECHT, G. A procedure for determining a useful approximation to an ideal scale. Unpublished manuscript.
13. SUCHMAN, E. A., & GUTTMAN, L. A solution to the problem of question bias. *Publ. Opin. Quart.*, 1947. (In Press.)
14. *Experiments on the measurement of the intensity function and zero point in attitude analysis*. Research Branch, Information and Education Division, Army Service Forces. Report D-1, Mimeographed, 1945.
15. *Questions and answers about scale analysis*. Research Branch, Information and Education Division, Army Service Forces. Report D-2, Mimeographed, 1945.
16. *The screening of neurotics*. Research Branch, Information and Education Division, Army Service Forces, Report B-104, 1944.
17. *A study of psychoneurotics in the army*. Research Branch, Information and Education Division, Army Service Forces. Report B-107, 1944.

NOTE ON "A REVIEW OF LEADERSHIP STUDIES WITH PARTICULAR REFERENCE TO MILITARY PROBLEMS"¹

DONALD E. BAIER

Personnel Research Section, A.G.O.

The valuable report² with which this note is concerned "... summarizes and reviews selected references from the available literature dealing with the problem of the selection of leaders in various fields. The primary interest in preparing the article was to provide a summary of techniques and results that would be of value to psychologists dealing with problems of selecting leaders, particularly in the military field."

It is the purpose of this note to make available additional facts and comments which appear to bear on the following conclusions of the reviewer:

1. "Progress has not been made in the development of criteria of leadership behavior"

2. "Advances in methodology in this field are definitely not striking."

It is this writer's belief that these conclusions, insofar as they are meant to apply to military leadership, are not entirely warranted.

In two reports³ published by the Medical Field Research Laboratory, Camp Le Jeune, N. C., research on measurement of "leadership" is reported. These studies indicate a substantial relationship (tetrachoric $r = .42$) between superior officers' reports of the combat performance of Marine Corps officers graduated from the Corps Officer Candidate School and the standing of these graduates among their fellow-marines as indicated by a nomination procedure conducted during their pre-officer training. The two sets of evaluations were completely independent.

An as yet unpublished follow-up study by the Personnel Research Section, AGO, of West Point graduates after 18 months of duty as Army officers also reveals a significant association ($r = .51$ for Infantry Officers) between inter-cadet ratings or leader-nominations and success as an officer measured by the *Officer Efficiency Report, WD, AGO Form 67*. Here again there is basis for believing that the two measures are independent.

¹ The opinions expressed herein are those of the author and do not necessarily represent the official view of the War Department.

² JENKINS, WILLIAM O. A review of leadership studies with particular reference to military problems. *Psychol. Bull.*, 1947, **44**, 54-79.

³ *Validation of officer selection tests by means of combat proficiency ratings*. Medical Field Research Laboratory Report No. 1, January 18, 1946 and No. 2, May 16, 1946. Camp Le Jeune, N. C.

The reviewer's account of the research upon which are based the current methods for selecting wartime officers for integration into the regular Army may result in misunderstanding. In discussing the correlation between the Officer Evaluation Report and the criterion of leadership, the latter being a product of nominations by subordinates and peers with a veto power resting with the commanding officer of the group, Jenkins states:

... The degree to which the Commanding Officers' ratings were weighted in the Officer Evaluation Report was not stated, but it appears likely that this factor played an important role. Substantial agreement between ratings by the C.O. and by fellow officers was to be expected. Since the OER had the highest validity, and the other measures when combined with it increased its correlation with the criterion only .07, these questions suggest the necessity for a further examination of the nature of the criterion here employed (p. 74).

The Officer Evaluation Report was accomplished in the majority of cases by the immediate supervisor, not the C.O., and represented only the former's evaluation of the ratee. The conclusion that substantial agreement between ratings by the C.O. and by fellow officers was to be expected does not appear to be justified. The C.O. was only one of from 7 to 30 nominators who participated in determining the ratee's criterion standing. He had no knowledge of how the other members of the nomination group evaluated each ratee, and his rating was used only to eliminate from the criterion groups of High, Low or Middle those rare cases where the C.O. placed the rated officer in the opposite extreme from the combined ratings of his subordinates and peers. Later studies employing a criterion which did not include the C.O. showed no drop in the validity of the OER or FCL type rating device. It is our belief that the nomination criterion employed in the studies cited does represent progress in the development of leadership criteria.

With respect to methodology, the forced choice technique as exemplified in the triads and tetrads of the OER and the recently revised Army Officer Efficiency Report seems to deserve more attention than the reviewer accords it. This technique, which has been described briefly in a paper titled "The Forced Choice Technique and Rating Scales," presented at the American Psychological Association meeting in Philadelphia on 5 Sept. 1946 by the Personnel Research Section, AGO, not only provides valid indicators of the ratees' standing on a nomination criterion, but favorably influences ratings of overall competence (if they are made immediately following completion of the FCL items) so that they show substantially less negative skewness. Clearly, the forced-choice technique is effective in diminishing rater-bias and in improving the distribution and validity of ratings which are generally regarded as indicative of leadership performance.

BOOK REVIEWS

MUNN, NORMAN L. *Psychology: The fundamentals of human adjustment*. Boston: Houghton Mifflin, 1946. Pp. xviii + 497.

The importance of the introductory course in psychology cannot be overestimated for it determines to a great extent the student's attitude toward the subject and whether or not he goes any further with it. But the importance of the textbook depends to a large degree upon the instructor. Some instructors lean heavily upon the text, others hardly at all. In reviewing a book, however, evaluation must be made as if it were the sole source of the student's introduction to the subject, regardless of the instructor's predilections, interpretations, choice of material, or method of handling the course. Though there are suggested readings at the end of each chapter in this as in other texts which the student is urged to consult, their influence is admittedly minor since the author of a text, as Munn says, writes with the feeling that he, in common with most teachers of the subject, could "organize its topics in a more logical sequence, choose apter illustrations, find more interesting examples and . . . write a book that . . . would be more appealing to instructors and students than any he has seen" (ii). Some other requirements of a good introductory text are succinctly suggested by President Leonard Carmichael in his Introduction wherein he discusses the reasons for studying Psychology today: as an essential part of a general education; as preparation for professions like law, medicine, teaching, the ministry, and business; and for further professional work in the subject itself. That Munn has met both his own and the editor's demands with considerable success there can be no question. The book is plentifully provided with excellent diagrams, half-tones, and tabular matter; it is full of concrete material chosen from a wide variety of sources; and its approach is scientific throughout.

It would seem, in the light of these virtues, that this text meets the requirements of the introductory course almost to perfection. It is such a splendid job in so many respects that any criticism at all seems supererogatory, if not hypercritical; and yet there are qualities expected in an elementary text which are of equal or greater importance than the ones which this book has in such large measure. The most important lack is an underlying point of view or theoretical structure integrating and unifying the topics and their relations to each other and the subject as a whole. We shall find that in this respect the book is not up to its accomplishments otherwise.

The book is divided into seven main parts and these in turn into two or more chapters. It begins with a discussion of general, methodological and historical material, then proceeds to consider in turn the anatomical and physiological bases of behavior, learning, remembering, thinking,

motivation, conflict, feeling, perceiving, the special senses, statistics, intelligence, aptitudes, and personality. The treatment develops by consideration of simpler processes followed by the more complex, in so far as possible, though there is some back-tracking which is done with a minimum of repetition of earlier material. With the general plan of the work before us we can now consider it more in detail.

Beginning with the origin and scope of psychology, the first two chapters are devoted to a brief glance at the history of the subject through a consideration of such topics as the psyche, the organism, methods in philosophy, physiology, and physics, analysis of consciousness and some fields of psychology. Chapters 1 and 2 really constitute a single topic or set of topics and furnish an excellent survey of methods, fields, and problems. They are properly brief, to the point, and very readable. Only in one detail does the text here need emendation. In discussing scientific controls it is stated that "there is never more than one independent variable in a given experiment If two or more factors were varied, he (the scientist) obviously would not know which had produced the phenomena observed" (p. 23). While it is not expected that the logic of analysis of variance and designed experiments should be presented at the elementary level (though it is not impossible) advances in statistics have made the older Mill-Bacon canons of scientific procedure represented by this statement quite out-of-date. Variation of only a single variable in psychological experiments is possible so seldom as to be almost a fiction and now that we have the statistical tools for handling multiple variates we might as well give up the fiction.

Part 2 deals with psychological development and consists of three chapters: origin and psychological significance of response systems, conception to maturity, and factors in psychological growth. Here the biological bases of behavior are explained and the psychological processes most directly correlated with them are brought in. The result is that the simpler and more complex processes are more or less intermingled in these chapters as a partial list of the topics reveals: structure and functions of receptor and nervous systems, embryonic development, sensitivity, locomotion, prehension, language, gestures, writing, genes, heredity, environment, and maturation. The order is in general from simple to complex but there are some reversals. Thus one would expect a discussion of genes and embryological development before discussion of the nervous system but here it follows the latter. The reason for Munn's order is obvious to a reader of the book and a good one: discussion of the more elementary gene units links directly with problems of heredity, environment, maturation and growth. There is no discussion of nerve action potentials and the treatment of the autonomic nervous system is postponed to the chapter on emotion where diagrams illustrating its relations to the cerebro-spinal system are given. The reviewer finds it impossible to omit the autonomic nervous system when

explaining the rest of the nervous system, although like Munn, he finds greatest use for it in the discussion of feelings and emotions. Figure 13 in chapter 3, showing the spinal reflex-arc system, would not be over-complicated if the sympathetic ganglion and its fibers were included as is usually done, and with some textual discussion this would remedy a serious omission at this point. On the other hand, Munn has included more material on the nervous system than is generally presented. The diagrams showing different types of synaptic connections make interaction, facilitation, and inhibition intelligible neurologically. The discussion of cortical representation of sensory functions is especially well done.

The chapters on conditioning, learning, memory and thinking succeed in presenting a considerable amount of material but suffer from the lack of clear integrating principles. While Munn rejects classical, Pavlovian conditioning theory as an adequate account of all learned behavior it is not clear what principles he would employ instead. That the author places greatest reliance on trial and error, past experience, and association appears from his treatment of certain particular problems rather than from explicit structuration of the material. One must dig his underlying approach out from a few critical cases which reveal the author's fundamental position. Thus the explanation of how the chimpanzee reaches an apparently inaccessible object is a case in point. According to Munn we are to suppose "... a chimpanzee has, in the jungle, learned to reach otherwise inaccessible objects by swinging toward them on a vine. Now in the psychological laboratory, he is confronted with an apparently inaccessible banana. A rope, however, is hanging nearby. *If the animal sees the similarity between the rope and the vine, or between his jungle method and the one now possible, he may solve the problem immediately*" (p. 122). (Italics are the reviewer's.) Now this explanation of the ape's accomplishment in terms of past experience which at first sight seems to be the scientifically simplest explanation actually turns out on closer scrutiny to demand much more in the way of memory and intellectual ability than the proposition that the animal simply sees the relevance of the rope to the banana which is immediately given. This assumption should not be too difficult to make since it was pointed out earlier that the difference between classical (mechanical) and instrumental conditioning lies in the fact that conditioning occurs much more easily when in the direction of more relevant responses. Certainly if the principle of relevance is basic to conditioning it may be accepted for the much more complicated case of insightful behavior. The welter of factors involved in acquiring skill and learning could be better ordered and made more meaningfully connected if some structure were seen behind the facts in question.

The lack of an adequate theoretical framework plagues the reader most in the concluding chapter on thinking. Reasoning, we are told, is

implicit trial and error, it is a form of controlled association, it is a combining of past experiences in order to solve problems which cannot be solved by mere reproduction of earlier solutions. At the same time the role of *direction* in reasoning and recall is emphasized, but how this factor operates with trial and error, past associations, and mere reproduction of earlier solutions to problems is not faced. We are here smack up against the problem of organization which many workers in biology as well as psychology realize cannot be dealt with adequately except as a problem in its own right. From the reviewer's experience such problems cannot be evaded even in the beginning course because many students have already faced them in courses in philosophy, logic, biology, and elsewhere.

The section on motivation of behavior seems to this reviewer to be best in the opening chapter dealing with physiological drives such as hunger, thirst, and sex where the material is largely drawn from experimental sources. The chapter on common social motives reads too much like a re-wording of the instinct psychology with too little use made of laboratory findings relevant to the topic. The chapter on conflict opens with sources of conflict in the environment and in the individual and then presents topological representation of conflict situations as an "interesting and illuminating method of representing and analyzing conflict situations" (p. 245). However, in the succeeding treatment of reactions to conflict such as compensation, identification, phantasy, projection, repression, and experimentally produced conflicts there is no further use made of Lewinian concepts. Again the integration must either be made by the instructor or the student will suffer from intellectual indigestion. Other alternatives are to omit topological representation or to put it at the very end of the chapter, pointing out that some, much, or most of the material discussed (depending upon the degree to which the instructor knows topological psychology) can be diagrammed in these terms. The author's penchant for trial and error pops out again in his recommendation of it as a possible solution in the alleviation or cure of conflict, contrary to the usual emphasis on rational procedures in psychotherapy. Since the patient admittedly knows why he is trying various lines of action, namely to find a way out of his conflict, it is doubtful if the procedure recommended is truly trial and error as Munn says.

The section on feeling and emotion which follows the one on motivation of behavior might well have preceded it, as affective states have been regarded by almost everyone as motivators of behavior. The main findings in the field are well covered with one or two exceptions. In the discussion of the Cannon-Bard theory the inhibitory function of the cortex is not mentioned and in the diagram illustrating the contrasting features of this theory as against the James-Lange theory the cortico-thalamic inhibitory path is not even shown. In view of the great im-

portance of the role played by the cortex in inhibiting emotion through positive inhibitory regulation and in allowing emotional expression through release of inhibition, the account offered here is entirely inadequate if not misleading, as reference to Bard's exposition in the *Handbook of General Experimental Psychology*, pp. 305-307, will show. Both text and diagram need this aspect of the theory for a correct as well as complete statement.

The following section, *Knowing Our World*, deals with attention and the special senses. Munn has here done an excellent job of boiling down the classical, and for the most part, stereotyped material and he has made it attractive by the use of well-chosen diagrams and half-tones. In view of the tremendous use made during the war years and now of material from the fields of sensation, perception, and psychophysics, not to mention their interrelations with sensori-motor learning, the time is past when we can rest content with traditional accounts of these fields. There is a wealth of material not yet in any text which modifies the whole approach to sensory processes and bears on every other field of psychology which should form part of the elementary student's equipment. Why recent work in some fields finds its way almost at once into textbooks and equally important work in other fields must wait a generation or more is hard to understand. For example, the explanation given of constancy is naive in the extreme in the light of not-so-recent work. And the epoch-making contributions by Katz find no reflection in the treatment of vision even though they had been in print for 35 years at the time this book appeared.

Several inaccuracies in terminology and fact should be corrected in future editions, such as: brightness and lightness should not be used interchangeably, and unless film and aperture modes are distinguished it is impossible to appreciate their difference; the assumption that Hering "neutral gray" is a constant or a general phenomenon rather than a special case is not tenable in view of work by Koffka and others; the discussion of retinal mixture versus overlapping of lights is so unclear it is impossible to determine what is meant and if it is correct; the usual explanation of *Flori* contrast as due to softening or obliteration of contours is palpably wrong and needs to be supplanted by the correct explanation given by von Bezold in the early part of the present century; the interchangeable use of note and tone, so common in discussions of hearing, should be replaced by more precise terminology in which tone relates to hearing-experience and note to the printed symbol. Only one figure of sensory qualities, the double cone for vision, is given although the smell prism and the taste tetrahedron are just as good in their respective modalities. The value of $1/3$ as the Weber fraction for temperature is altogether too large to be representative following Culler's work.

In general the chapters just considered rely too much, on the theoretical side, on past experience and similar explanations and suffer

from a lack of unifying principles by which intra- and inter-sensory material may be related as well as unified with other psychological processes. If, as admitted, principles of organization are effective in attending, are they not perhaps also of importance in perception, and taking a further step, in learning and thinking as well? Recognition of such principles might unify and simplify psychology for the beginner.

The seventh and final section of the book, *Individual Differences*, contains chapters on statistics, intelligence, aptitudes, and personality. The chapter on statistics, kept until this section as an "Introduction to Statistical Analysis of Individual Differences," might well come earlier, especially for use in courses with laboratory. However the chapter can be introduced as it stands in almost any part of the course so its actual position matters little. The other three chapters form a fitting close to the book, entirely in the spirit of the more experimental portions in being packed with concrete material. Intelligence is approached from the historical angle and the important question of heredity and environment is quite fully discussed. The discussion of factor analysis—including the fundamental factors found by Thurstone, and the illustratory material from test batteries—make this chapter unique for an elementary presentation and one of the finest things in the book.

Similarly the chapters on aptitudes and personality are extremely well presented and again demonstrate the author's ability to condense a large amount of fact into a relatively small compass. In the chapter on personality the discussion includes methods of approach such as case history, rating, paper and pencil tests, behavior tests, interviews, free association, dream analysis, and projective methods, and also physique and temperament, role of the endocrines, and abnormal states. The open, empirical treatment here is more acceptable because the subject is more familiar to the average student and personality as a concept already provides some structuration by which its data can be ordered.

Taking the book as a whole, what are its pros and cons? On the plus side it is an excellent text in so far as it provides a wealth of concrete factual data chosen from widely different sources both within and outside psychology proper. With some exceptions it represents present-day scientific psychology very fairly. The student should come away from this text with respect for the scientific approach not because he has been told that it works best in various fields but because he has found that material obtained by scientific methods can be applied to many different life situations and leads to further fruitful discovery. If the author is unable to accept a theory *in toto* his criticism is so mild and fair that the student's respect for the theory as well as for psychology in general is in no wise diminished. This and the catholicity of Munn's approach should exert a very good effect on coming generations of psychologists. Too often personal or institutional loyalties lead even graduate students to belittle men and work done outside their own bailiwicks with effects

detrimental both to themselves and to psychology. This book should serve as an excellent corrective to this tendency.

On the negative side of an otherwise fine piece of writing and presentation must be noted the lack of an integrating and unifying point of view which has been pointed out in our previous discussion. This lack results in a looser and more disjointed treatment than is necessary in the light of present advances along various fronts. This is not meant to imply that Munn himself does not have a point of view. As we have seen, careful reading reveals that for him trial and error is the great principle operating in human behavior and a number of indications are present that he believes in what has been dubbed an "atomistic logic," *i.e.*, proceeding from "simples" to "complexes." But having brought into his discussion of conditioning the principle of relevance, into thinking the principle of direction, into some sensory experiences primitive organization, and having recognized other whole-properties as well, he is under obligation to apply them more generally where they are applicable or at least to square them with the fundamental principles he believes are operative. Perhaps he has done this and this reviewer has missed it. If so, then it is probable that most students will fail to see how it all goes together.

The tendency to over-simplify has already been pointed out with reference to certain neural diagrams but it occurs much more frequently in the textual discussion where it leads the author, in making points which are quite valid in themselves, to say things he cannot possibly mean as they stand. For example, in writing about the influence of animal experiments on the theoretical basis of psychology, there appears the remarkable statement: "After all, learning is learning and vision is vision whether it occurs in man or animal" (p. 10). But the vision of the most widely used laboratory animal, the white rat, is very different from that of man, from retina to higher cortical centers, and Munn later points out that "Insight is rare in animals, not quite so rare in children, and quite common in human adults" (p. 109), meaning to distinguish among kinds of learning. One finds too many statements like this which take a good deal of explaining to mitigate.

There can be little doubt the present book will set a pattern for future introductory texts. The double columns while providing a shorter reading line and more words per page also make possible wider spaces for illustrative material and marginal notes. The wealth of charts, diagrams, and pictures lessens the instructor's blackboard work and should prove a boon to places where laboratory work cannot be given. In this text psychology appears as a positive, if not positivistic, science. If it were possible to combine what Munn has done with more emphasis on methods and unifying principles we should be much nearer the perfect presentation of present-day psychology everyone desires.

HARRY HELSON.

BRIDGES, J. W. *Psychology normal and abnormal*. Toronto: Sir Isaac Pitman & Sons, 1946. Pp. xviii+470.

Except for dropping the chapter on philosophical foundations, the splitting and partial revision of the chapter on reflexes and instincts, and the annotation of the extensive bibliography, the 1946 edition has "identical twin" resemblance to its 1930 predecessor. The appearance of the revision does, however, call attention again to a book with a classical timelessness of integration (despite an eclectic tolerance), stimulating hypotheses, and a style reminiscent of William James. The general reader and even the professional psychologist will find interest and value here, although the book was designed for the introductory psychology course of pre-medical and medical students.

Bridges chooses to give the distilled essence of a topic rather than to lead the reader to a conclusion from the raw data of experiments and case studies. The few graphs, tables, and other reference to specific studies are illustrative only. Just as the dramatist's words furnished the bare Elizabethan stage, the emphasis on the logic of the argument in this book seems to stimulate more associations and imagery than one gets from many texts replete with illustrations. Many more of the quotations are from English and French psychologists than one meets in most American texts.

The chapter headings might have come from any of a number of general psychologies, but the plan of devoting the first half of each chapter to normal behavior and the remainder to the related abnormalities makes a distinctive pattern throughout the book. Technical words are italicized and well defined. The chapter on applied psychology delimits that field with such precision and perspective that it ought to be widely reprinted.

An error not corrected from the 1930 edition is the taking of the standard deviation from the median. Emphasis on the older studies, e.g., Downey will-temperament tests, is heavier than on those of the last two decades.

The problem of how to teach psychology in medical schools or to pre-medical students seems to have led in at least three main directions. Some have emphasized a sociological-psychological approach, as in Pressey's *Life*, since this is a common medical blind spot. Others have stressed the genetic-psychosomatic attitude, as found in such authors as Maslow and Mittlemann. A third group believe that medical students are more highly motivated and gain more insight from the contrasts and comparisons of the normal and abnormal. Bridges from his experience as the first professor of abnormal psychology on a medical faculty has provided an effective text for the last group.

GEORGE M. HASLERUD.

University of New Hampshire.

GRAY, J. S. *Psychology in human affairs*. New York: McGraw-Hill, 1946. Pp. viii+646.

While this book is, in many respects, a successor to Gray's previously edited text, *Psychology in Use* (American Book Co., 1941), in that it discusses the applications of psychology to the main fields of practical life, and represents the co-authorship of eleven other contributors, nevertheless it is not merely a revision. With two exceptions, the co-authors are new. They are less well known than those of the former book, but Gray has himself taken a more active part in the actual writing of the text. Several chapters appear for the first time, such as "Psychology in Speech Correction," "Psychology in Music, Art, and Literature," and "Psychology in Military Affairs." Others appear under new titles, and are written from a new viewpoint.

Perhaps the outstanding characteristic of the book is its emphasis on factual material. For example, Chapter II, on "Psychology in College Life" contains twenty tables and four graphs. Chapter III, on "Child Development" contains ten graphs and twenty-one tables. Much of this material is new to textbooks, and with few exceptions, the references are to studies published after 1930. The general effect of this emphasis on experimental data and practical findings is to require a change in teaching methods on the instructor's part. His function is no longer to supplement the text with up-to-date illustrative material, but rather to interpret and evaluate that which is given. Less supplementary assigned reading is needed, and much more digesting of the text by the student. The art of reading and interpreting tables and graphs is one which requires special training. Many students are allergic to statistics, though this is not in itself an argument for using them sparingly, if the instructor is competent to vitalize them. But the chief value of facts is to illustrate and support laws, principles and theoretical formulations. They are most effectively used in the inductive development of a topic. Most of these facts are and should be promptly forgotten by the student, so that his memory is freed for the permanent retention of the principles. The immature student needs much expert guidance in recognizing the bare essentials of fact to be learned. While this book is many strides ahead of the type which presents only unsupported assertions, or illustrations selected for their patness only, does it err slightly in the opposite direction?

Another important characteristic of the book is its emphasis on the practical. This is to be expected in an applied psychology text, but is seldom achieved. Omnibus books too often give an impression of sketchy remoteness, with little practical contact, while technical treatises are written for the advanced student who wants specialized information. This book, by achieving a compromise between these two extremes and by emphasizing the practical aspects of each field for the layman, fills a

real need. Its range of topics is wider, its treatment more complete than is customary in such books.

In spite of its up-to-dateness, the book occasionally presents old data which have been superseded, or theory which is now modified. In one or two cases, quite erroneous statements appear, such as the following, in connection with a discussion of the topic of I.Q. constancy, on pages 91-92:

If the child develops mentally at exactly the same rate as other children tested, his I.Q. will remain constant. However, if his mental development is faster than that of other children, his I.Q. will increase. Likewise, if his mental development is slower than that of other children, his I.Q. will decrease.

The author seems to have confused *constancy* of I.Q. with *normality* of I.Q., for if the statement were taken as it stands, it would mean that no child's I.Q. is constant if he has a faster or slower developmental rate than the average child.

An innovation in this book which probably has pedagogical value and would be used oftener by authors of texts if publishers would let them, is the table of contents at the beginning of each chapter. The addition of page numbers would increase its value. A word must be said in criticism of certain reproduced charts, in which the reduction in size of print needed to get them on the page has made them unreadable; for example, those on pp. 474 and 573. Otherwise the style of the book is good.

Many psychology teachers will welcome this book either as a supplement to the general course in psychology, or as a second course to follow the introductory one. Those who teach adult extension classes will find it an excellent survey text, both meaty and comprehensive.

ARTHUR G. BILLS.

University of Cincinnati.

LUCK, J. M., & HALL, V. E. (Eds.). *Annual review of physiology* (Vols. VII & VIII). Stanford Univ. P.O.: Annual Reviews, Inc. and American Physiological Society, 1945, 1946. Pp. vi+774 (Vol. VII). Pp. vi+658 (Vol. VIII).

These are Volumes VII and VIII of the annual series begun in 1937 and published jointly by the American Physiological Society and Annual Reviews. It is the declared editorial policy of the *Review* that "encouragement is given only to preparation of reviews which survey the important contributions of the preceding year or biennium, which appraise them critically, and evaluate with discrimination the present status of the subject. Comprehensive reviews in which the task of the author is one of compilation rather than of appraisal are deliberately eschewed." Despite this policy, some of the reviews are principally compilations or annotated bibliographies. And some are rather spotty

compilations mixed with evaluation, while only a few reach the goal of really critical reviews of the literature. On the whole, however, the reader can obtain a picture of the more significant aspects of research progress in the respective fields covered by the review.

Volume VII contains 26 chapters written by 30 authors, and Volume VIII contains 25 chapters by a total of 29 authors. In each case, bibliographies of literature cited in the various reviews total about 4,000 references. At the end of each volume is an author index of about 4,000 items and a subject index of about 40 pages in length.

Because the reviews are written by physiologists for physiologists, more than half of each volume is of no interest to the psychologist except that occasionally there is a brief treatment in a sentence or two of some psycho-physiological problem. The psychologist, however, who wants to find out what has happened recently in some special aspect of physiology relevant to his field of teaching or research will find that his best bet is to go to these volumes and to consult the excellent author and subject indices before resorting to other less up-to-date textbooks or to more laborious methods of library research.

More than that, however, many of the special chapters in physiology are good reading for psychologists engaged in the respectively related field of psychology: for genetic psychology, *Physiological aspects of genetics* (VII and VIII) and *Developmental physiology* (VII and VIII); for sensory psychology, *The special senses* (VII) and *Audition* (VIII); for neural mechanisms of behavior, *Electrical activity of the brain* (VII), *Conduction and synaptic transmission in the nervous system* (VII), *Nerve and synaptic conduction* (VIII), *The visceral functions of the nervous system* (VII and VIII); and for a general review of physiological psychology, *Physiological psychology* (VII and VIII).

The contrast between the chapters on physiological psychology in the two volumes deserves a special comment. In Volume VII, Stone has presented a careful and critical review. He has covered thoroughly the recent literature and has appraised its strength and shortcomings so that the reader can see what has happened and what it means for physiological psychology.

The same chapter in Volume VIII by Seashore, however, does neither of these two things. It opens with a philosophical discussion of the mind-body problem and of the scientific approach to it. The chapter then proceeds to summarize the present status of individual differences in skills, abilities, aptitudes and capacities. Finally, it gives a general summary of the effects of extreme working conditions upon the effectiveness of human performance. Thus Seashore's chapter spends a lot of time on problems which are not physiological psychology, in any reasonable definition of the field; and he fails to review or appraise the recent literature in the field. As a consequence, the physiologist reading the two chapters is likely to be bewildered by two so very different concepts of physiological psychology.

Looked at in perspective, these two volumes of *Annual Review of Physiology*, like previous volumes in the series, are an extremely valuable aid, not only to physiologists, but to all those for whom physiology is an important ancillary subject. By and large, the chapters give scholarly up-to-date appraisals of their respective fields. As he has stated before, this reviewer feels that a companion volume, giving an annual review of psychology, would be an invaluable aid to psychologists, which would help us "keep up with the literature" and give us better perspective on the developments in our field.

CLIFFORD T. MORGAN.

The Johns Hopkins University.

BARKER, ROGER G., WRIGHT, BEATRICE A., AND GONICK, MOLLIE R. *Adjustment to physical handicap and illness: a survey of the social psychology of physique and disability*. New York: Social Science Research Council, 1946. (Bulletin No. 55.) Pp. xi+372.

This is another in the excellent series of research summaries sponsored by the Social Science Research Council. The authors have earned special commendation by providing intelligently critical comments as to the assumptions and thinking of earlier investigators, rather than merely reporting data and conclusions; by writing this summary of prior research into a theoretical frame of reference (topological psychology); and by introducing some well-chosen original material to illuminate the inferences drawn from published sources.

From their survey the authors have eliminated somato-psychological studies of age, sex, race, and speech defects, on the ground that these have recently been covered adequately by other reviewers. Leprosy is discarded as a minor problem in the western world. Of the remaining areas, detailed reports are presented on: normal variations in physique; crippling; tubercular conditions; impaired hearing; and acute illness. Bibliographies are added on: visual disability; cardiac conditions; diabetes mellitus; cosmetic defect; rheumatism; and cancer.

The least satisfactory chapter is that on variations in normal physique. There is a good section on size changes at adolescence, but the discussion of variations in adult size is rather elementary, and no mention whatever is made of Sheldon's *The varieties of human physique* and *The varieties of temperament*. Even if one does not accept Sheldon's theory, his work can hardly be ignored. The authors occasionally dip a hesitant toe into the cold waters of endocrinology, genetics and autonomic nervous function, then withdraw hastily. Clarity would have been improved by frankly excluding such material.

Outstanding treatments of crippling, tubercular conditions and impaired hearing more than atone for any shortcomings of the earlier chapter. Particularly interesting is the mode of analysis in terms of *overlapping situations*. A disabled person is able to function on a par with normals in some environments; he is decisively barred from other

situations; but between these extremes will fall a range of ambiguous conditions in which the handicapped person may participate, but under difficulties. The Lewinian concepts of barrier, valence, potency and congruence are used fruitfully to show basic similarities between the situations facing the orthopedic cripple, the tubercular, the deaf and the individual with acute illness.

The person with impaired hearing, for example, functions in many situations unnoticed by his normal associates. If the behavior involved does not require auditory controls, he may compete on equal terms. Where hearing is involved, he may be handicapped and subject to extra criticism, since his impairment is not obvious and many normals (e.g., school teachers) fail to make allowances for it. The barriers in his field are indefinite (as contrasted with the orthopedic cripple, for example, to whom certain activities are plainly impossible), and this condition often gives rise to vacillation and instability. The valence of full-normal activity is positive and high, but the valence of failure and criticism is negative and high. Thus physically handicapped persons are likely to show the familiar symptoms of conflict.

We occasionally feel, in these topological analyses, that there is an unstated shift from the topology of the external situation (geographical environment) to the situation as perceived by the individual (behavioral environment). In the case of cripples, for example, some activities are objectively impossible, whereas others are subjectively considered to be impossible. It is clear that these two should not be treated as identical, and yet that impression is sometimes given. If the entire analysis were erected on a perceptual basis, this uncertainty could have been avoided.

The importance of the individual's perception of his defect, and of his behavioral field, is well illustrated by the discussion of family attitudes toward the handicapped. Many parents reject the handicapped child, while others over-protect him and keep him in an infantile status. Optimum adjustment seems to be achieved when the parents adopt an understanding, objective attitude which focuses the child's attention on realistic assessment of the situation. Excessive sympathy and pity are likely to encourage exaggeration of barriers and exclusion of many possibilities for normal participation.

The final chapter on employment of disabled persons gives a realistic and well-considered treatment of this problem, the solution of which is basic to optimum adjustment of handicapped adults.

ROSS STAGNER.

Dartmouth College.

LEWIS, CLAUDIA. *Children of the Cumberland*. New York: Columbia Univ. Press, 1946. Pp. xviii+217.

Before going to the Southern highlands, Miss Lewis was a teacher in the Harriet Johnson Nursery School in Greenwich Village, New York

City. She compares the behavior of the children in the nursery school which she established in the mountains of Tennessee with the behavior of her Greenwich Village pupils. The majority of the material presented in the book concerns the mountaineer subjects.

A considerable part of the volume consists of a collection of incidents involving child care or child behavior. These range in length from single sentences to a page or more, in age of subjects from birth to senility, in form from dialogue to descriptive essays. They are extremely readable and serve to render very vivid the life of the mountain people.

Miss Lewis does not attempt to present quantitative measures, but for this she cannot be reprimanded. It is apparent that she devoted very full days to the nursery school, and that research had a secondary place. Nevertheless, her thinking is quantitative. She emphasizes the diversity of individual behavior which takes place in both schools, and makes clear that there is overlapping between the schools. However, it is her belief that there are large differences in central tendencies between her two kinds of subjects. With this contention, it is likely that nearly every person who is familiar with both types of children will agree.

Miss Lewis finds more spontaneity, more energy, more conflict, more aggression in the New York group. The mountain children are more placid, more compliant, more quiet, and in some respects, better adjusted.

She does not find the explanation of these differences in any one factor. Among the probable causes mentioned are the following: the differences in spaciousness of the environment, differences in climate, health, and nutrition, differences in sleeping habits, differences in infant care and family structure, differences in discipline, and differences in environmental stimulation.

Throughout the book, Miss Lewis shows an excellent understanding of child development in both New York and Tennessee—not an easy achievement. She also displays a high degree of ability as a writer. This combination of traits makes her book one of the best in the "child in a culture" field. It should be profitable alike to teacher, parent, psychologist, and sociologist.

The reader may sample for himself some of Miss Lewis's attitudes and style in the following quotation from her concluding chapter:

No, now that this study is made I am not packing my trunks with the intent of moving down to Tennessee, building me a cabin, taking to the "simple life" and rearing my hypothetical children in the way the Summerville families do. For us it is not a question of attempting to turn the clock back in that way, which, indeed, would be as impossible as it would be undesirable. It is rather a question of trying to bring to Greenwich Village a little more of . . . Summerville life . . .

WAYNE DENNIS.

University of Pittsburgh.

situations; but between these extremes will fall a range of ambiguous conditions in which the handicapped person may participate, but under difficulties. The Lewinian concepts of barrier, valence, potency and congruence are used fruitfully to show basic similarities between the situations facing the orthopedic cripple, the tubercular, the deaf and the individual with acute illness.

The person with impaired hearing, for example, functions in many situations unnoticed by his normal associates. If the behavior involved does not require auditory controls, he may compete on equal terms. Where hearing is involved, he may be handicapped and subject to extra criticism, since his impairment is not obvious and many normals (e.g., school teachers) fail to make allowances for it. The barriers in his field are indefinite (as contrasted with the orthopedic cripple, for example, to whom certain activities are plainly impossible), and this condition often gives rise to vacillation and instability. The valence of full-normal activity is positive and high, but the valence of failure and criticism is negative and high. Thus physically handicapped persons are likely to show the familiar symptoms of conflict.

We occasionally feel, in these topological analyses, that there is an unstated shift from the topology of the external situation (geographical environment) to the situation as perceived by the individual (behavioral environment). In the case of cripples, for example, some activities are objectively impossible, whereas others are subjectively considered to be impossible. It is clear that these two should not be treated as identical, and yet that impression is sometimes given. If the entire analysis were erected on a perceptual basis, this uncertainty could have been avoided.

The importance of the individual's perception of his defect, and of his behavioral field, is well illustrated by the discussion of family attitudes toward the handicapped. Many parents reject the handicapped child, while others over-protect him and keep him in an infantile status. Optimum adjustment seems to be achieved when the parents adopt an understanding, objective attitude which focuses the child's attention on realistic assessment of the situation. Excessive sympathy and pity are likely to encourage exaggeration of barriers and exclusion of many possibilities for normal participation.

The final chapter on employment of disabled persons gives a realistic and well-considered treatment of this problem, the solution of which is basic to optimum adjustment of handicapped adults.

ROSS STAGNER.

Dartmouth College.

LEWIS, CLAUDIA. *Children of the Cumberland*. New York: Columbia Univ. Press, 1946. Pp. xviii+217.

Before going to the Southern highlands, Miss Lewis was a teacher in the Harriet Johnson Nursery School in Greenwich Village, New York

City. She compares the behavior of the children in the nursery school which she established in the mountains of Tennessee with the behavior of her Greenwich Village pupils. The majority of the material presented in the book concerns the mountaineer subjects.

A considerable part of the volume consists of a collection of incidents involving child care or child behavior. These range in length from single sentences to a page or more, in age of subjects from birth to senility, in form from dialogue to descriptive essays. They are extremely readable and serve to render very vivid the life of the mountain people.

Miss Lewis does not attempt to present quantitative measures, but for this she cannot be reprimanded. It is apparent that she devoted very full days to the nursery school, and that research had a secondary place. Nevertheless, her thinking is quantitative. She emphasizes the diversity of individual behavior which takes place in both schools, and makes clear that there is overlapping between the schools. However, it is her belief that there are large differences in central tendencies between her two kinds of subjects. With this contention, it is likely that nearly every person who is familiar with both types of children will agree.

Miss Lewis finds more spontaneity, more energy, more conflict, more aggression in the New York group. The mountain children are more placid, more compliant, more quiet, and in some respects, better adjusted.

She does not find the explanation of these differences in any one factor. Among the probable causes mentioned are the following: the differences in spaciousness of the environment, differences in climate, health, and nutrition, differences in sleeping habits, differences in infant care and family structure, differences in discipline, and differences in environmental stimulation.

Throughout the book, Miss Lewis shows an excellent understanding of child development in both New York and Tennessee—not an easy achievement. She also displays a high degree of ability as a writer. This combination of traits makes her book one of the best in the "child in a culture" field. It should be profitable alike to teacher, parent, psychologist, and sociologist.

The reader may sample for himself some of Miss Lewis's attitudes and style in the following quotation from her concluding chapter:

No, now that this study is made I am not packing my trunks with the intent of moving down to Tennessee, building me a cabin, taking to the "simple life" and rearing my hypothetical children in the way the Summerville families do. For us it is not a question of attempting to turn the clock back in that way, which, indeed, would be as impossible as it would be undesirable. It is rather a question of trying to bring to Greenwich Village a little more of . . . Summerville life . . .

WAYNE DENNIS.

University of Pittsburgh.

LEEPER, ROBERT. *Psychology of personality*. Ann Arbor: J. W. Edwards, 1946. Pp. 167.

The format of this book, resembling that of a laboratory manual or workbook, may lead many to overlook this significant treatment of personality and mental hygiene.

The author's organization apparently proceeds from an assumption which has increasingly impressed itself on the reviewer in recent years: namely, that, in order to have functional significance, any treatment of mental hygiene must be based on a consistent theory of personality processes. Moreover, such a treatment should not be left among the author's un verbalized potentialities, but should be given systematic formulation. It is to this task that the major emphasis of Leeper's book is devoted.

Leeper's treatment is mainly an elaboration of the following thesis:

In general, the term "personality" covers three things: (1) the person's motives, and especially his emotional motives, or ways in which he responds emotionally in different life situations; (2) the general techniques by which, characteristically, he tries to attain satisfaction for these motives; and (3) the background of meanings or pictures of reality which determine the motives and types of adjustive responses of the person (p. 5).

The discussion of motivation distinguishes between physiological and emotional motives and between positive and negative emotional motives. While the latter distinction seems arbitrary, since the "negative" motives can be regarded as the products of frustration of the "positive" motives, it serves a useful expository purpose when the author deals with motivational differences existing between well adjusted and poorly adjusted personalities.

The techniques by which the person tries to attain satisfaction for his motives are considered from two points of view: (1) the nature of the learning processes, and (2) the description of effectual and ineffectual adjustment techniques. The learning processes are treated with due attention to the dynamic complexities recognized in modern learning theory. In addition to describing the usual techniques employed by maladjusted personalities, the author discusses some of the major techniques by which superior personalities distinguish themselves.

In his discussion of the "background of meanings," the author deals competently with an aspect of behavior which seldom receives the emphasis merited by its significance for personality dynamics. Leeper's thesis is that "... a person cannot govern his behavior just by what is objectively and actually true, but ... must forever live and react in terms of properties which he infers as existing because of his experience in previous situations" (p. 92). This view that behavior is determined not as much by the character of objective reality as by the individual's interpretation of reality (through the phenomenon of emotional transference) is supported in terms of the principle of equivalence of stimuli and the principle of substitute response or displacement. Treated in

these terms, the "background of meanings" is seen to be an aspect of personality whose importance has been emphasized by such widely separated disciplines as the research of animal psychologists and the clinical observations of psychoanalysts.

It is the reviewer's opinion that through the medium of Leeper's book the principles of personality functioning are made understandable to the average undergraduate without undue simplification or loss in organic quality. For this reason, it seems regrettable that the book was not produced in a form more likely to have wide distribution.

The book will probably be disappointing to those who feel that a textbook should serve as a compendium of psychological research findings. While the author draws freely upon research sources, these tend to lose their distinctive identities in the author's discussion. No bibliography or index is provided.

BERT R. SAPPENFIELD.

Montana State University.

KELLEY, DOUGLAS M. *22 cells in Nuremberg*. New York: Greenberg, 1947. Pp. 245. \$3.00.

The author of this book was for five months the official psychiatrist at the Nuremberg prison and in that capacity made psychiatric examinations of all the 22 top-ranking Nazi prisoners. The customary medical and psychiatric procedures were supplemented by Rorschach personality tests and Wechsler-Bellevue intelligence tests given by the author's fellow officer, Capt. G. M. Gilbert. The examinations and tests were further supplemented by information obtained from former intimate associates of the accused and from motion pictures, speeches, writings, and other records.

Except for Rudolf Hess and Hermann Goering, this was the first psychiatric study to be made of any of the accused. In view of the fact that twelve of the group are no longer living and that all but three of the others were disposed of by prison sentences of from ten years to life, the documents and conclusions of Dr. Kelley are destined to be of lasting historical interest.

The task which the author set himself was not merely or chiefly to determine the degree of mental responsibility of the subjects, but rather to investigate their basic personality patterns. He wanted to find out what these men were like who had made themselves masters of eighty million people, and what factors in childhood, youth, and later years had made them what they were. The book attempts to answer these questions in language sufficiently nontechnical to be intelligible to readers who are relatively unfamiliar with esoteric theories of personality. We are informed that more detailed reports of the work will be published later in professional journals, and that transcripts of the interviews and other records will ultimately be available to historians.

Among those to whom most space is devoted are Goering (27 pages), Hess (22 pages), Ley (21 pages), Rosenberg (13 pages), and Streicher

(11 pages). The other 17 members of the group get from three to eight pages each. By good luck the examination of Ley had been completed before he committed suicide, and we are told that the post-mortem examination of his brain confirmed the psychiatric diagnosis.

With the exception of one chapter, the book is concerned entirely with the 22 Nazis who were studied first-hand by the author. The additional chapter presents a 35-page portrait of Hitler based on information and comments obtained from the Führer's contemporaries, his aides, his personal physicians, and his secretaries. Some of this information is new, and the author's interpretations differ in several important respects from those which have been current.

Within the limits of a brief review it is not possible to summarize the author's interpretations of the individual subjects. In fact, each portrait as sketched is a unified gestalt that almost defies further condensation. The sitters for these portraits composed indeed a motley group. They ranged from the stupid to the highly intelligent; from the semi insane to the stable and well integrated; from the shrewd and talented leader to the errand-boy hanger-on seeking in Hitler a father surrogate. But there were three characteristics which they had in common: inordinate ambition, debased ethical standards, and a hyperdeveloped nationalism that justified anything done in the name of Germanism—plus, of course, an economic and political environment that allowed full play to their ruthless wills.

The author's conclusion is that Nazism was a "socio-cultural disease," epidemic among our enemies but endemic everywhere. He tells us that the Nazi leaders were not the rare and spectacular types that can be expected to appear only once in a century. Instead, neurotics like Hitler, with "hysterical disorders and obsessive complaints, can be found in any psychiatric clinic." Similar ones, thwarted and discouraged, but determined to do great deeds, roam the streets of every American city. "Strong, dominant, aggressive, and egocentric personalities like Goering . . . can be found anywhere—behind big desks deciding big affairs as businessmen, politicians, and racketeers." We hardly need to be reminded that men strongly resembling some of these types occasionally win election to our highest law-making bodies or to the governorship of a great state.

Dr. Kelley has analyzed for us 22 types of totalitarian-virus, has described the soil in which they thrive, and has indicated some of the means by which society can protect itself against them. His book will inevitably be compared with one written by another psychiatrist—Brickner's *Is Germany Incurable?* Of the two, the reviewer finds Kelley's less controversial and no less challenging.*

LEWIS M. TERMAN.

Stanford University.

* Since this review was written, *Nuremberg Diary*, by Capt. G. M. Gilbert, has been published. This book should be read along with Dr. Kelley's. L.M.T.

BOOKS AND MATERIALS RECEIVED

- ALLPORT, G. W., AND POSTMAN, LEO. *The psychology of rumor*. New York: Henry Holt, 1947. Pp. vii+247.
- ALSCHULER, ROSE H., AND HATTWICK, LA BERTA W. *Painting and personality*. (2 Vols.) Chicago: Univ. of Chicago Press, 1947. Pp. xi+590.
- AXLINE, VIRGINIA MAE. *Play therapy*. Boston: Houghton Mifflin, 1947. Pp. xii+379. \$3.50.
- BARUCH, DOROTHY, AND MONTGOMERY, ELIZABETH. *The girl next door*. Chicago: Scott, Foresman, 1947. Pp. 256.
- BAUMGARTEN-TRAMER, F. *Der Rorschach—Test im Lichte der Experimentellen Psychologie*. Archivio di Psicologia Neurologia e Psichiatria, 1946, 8, No. 2. Pp. 37.
- BECK, H. P. *Men who control our universities*. New York: King's Crown Press, 1947. Pp. x+229.
- BOWERMAN, WALTER G. *Studies in genius*. New York: Philosophical Library, 1947, Pp. 343.
- CARTER, L. F. (ED.) *Psychological research on navigator training*. Army Air Forces Aviation Psychology Program, Research Reports. Report No. 10. Washington: U. S. Government Printing Office, 1947. Pp. iii+186.
- COLE, LUELLA, AND MORGAN, J. J. B. *Psychology of childhood and adolescence*. New York: Rinehart, 1947. Pp. xi+416.
- COUNT, E. W. *Brain and body weight in man: their antecedents in growth and evolution*. Annals of the New York Academy of Sciences, 1947, Vol. XLVI, Art. 10. Pp. 993-1122.
- COWLES, E. S. *Don't be afraid*. Chicago: Wilcox & Follett, 1947. Pp. xv+254.
- DAVIS, F. B. (ED.) *The A. A. F. qualifying examination*. Army Air Forces Aviation Psychology Program, Research Reports. Report No. 6. Washington: U. S. Government Printing Office, 1947. Pp. iii+266.
- DAVIS, F. B. *Utilizing human talent*. Washington: American Council on Education, 1947. Pp. ix+85.
- DICHTER, E. *The psychology of everyday living*. New York: Barnes & Noble, 1947. Pp. x+239.
- ERICKSON, C. E. (ED.) *A basic text for guidance workers*. New York: Prentice-Hall, 1947. Pp. x+566.
- FALVEY, HAL. *Ten seconds that will change your life*. Chicago: Wilcox & Follett, 1947. Pp. 96.
- FIELD, HARRY. *Midiendo la Opinion Publica*. Mexico, D. F.: In-

stituto de Estudios de Psicología Social y Opinión Pública, 1945. Pp. 56.

FORD, MARY. *The application of the Rorschach test to young children*. Institute of Child Welfare Monograph, No. 23. Minneapolis: Univ. of Minnesota Press, 1946. Pp. xii+114.

FREDERICK, R. W., KITCHEN, P. C., AND MCELWEE, AGNES R. *A guide to college study*. New York: Appleton-Century, 1947. Pp. viii+341.

GARRETT, H. E. *Statistics in psychology and education*. (3rd Ed.) New York: Longmans, Green, 1947. Pp. xii+465.

GILBERT, G. M. *Nuremberg diary*. New York: Farrar, Straus, 1947. Pp. 471.

HAGAN, W. A., *et al.* *The relation of diseases in the lower animals to human welfare*. Annals of the New York Academy of Sciences, 1947, Vol. XLVIII, Art. 6. Pp. 351-576.

HALL, R. B. *Area studies: with special reference to their implications for research in the social sciences*. New York: Social Science Research Council, 1947. Pp. iii+90.

HALL, V. E. (Ed.) *Annual review of physiology* (Vol. IX, 1947). Stanford Univ. P.O.: Annual Reviews, Inc. & American Physiological Society, 1947. Pp. vii+736.

JELLINEK, E. M. *Recent trends in alcoholism and in alcohol consumption*. New Haven: Hillhouse Press, 1947. Pp. 42.

KITAY, P. M. *Radicalism and conservatism toward conventional religion. A psychological study based on a group of Jewish college students*. Teachers College, Columbia Univ., Contr. to Educ., No. 919. New York: Bureau of Publications, Teachers College, Columbia Univ. 1947. Pp. vii+117.

KRACAUER, S. *From Caligari to Hitler*. Princeton: Princeton Univ. Press, 1947. Pp. xi+361.

LE CRON, L. M., AND BORDEAUX, JEAN, *Hypnotism today*. New York: Grune & Stratton, 1947. Pp. v+278.

LEIGHTON, DOROTHEA, AND KLUCKHOHN, CLYDE. *Children of the people*. Cambridge: Harvard Univ. Press, 1947. Pp. viii+277.

LEPLEY, W. M. *Psychological research in the theaters of war*. Army Air Forces Aviation Psychology Program, Research Reports. Report No. 17. Washington: U. S. Government Printing Office, 1947. Pp. iii+201.

LINDNER, R. M., AND SELIGER, R. V. *Handbook of correctional psychology*. New York: Philosophical Library, 1947. Pp. 691.

LINK, H. C. *The rediscovery of morals*. New York: E. P. Dutton, 1947. Pp. 223.

LONDON, L. S. *Libido and delusion*. (2nd Ed.) Washington: Mental Therapy Publ., 1946. Pp. xi+259. \$3.50.

LOUTTIT, C. M. *Clinical psychology*. (Rev. Ed.) New York: Harper, 1947. Pp. xviii+661.

LUNDBERG, G. A. *Can science save us?* New York: Longmans, Green, 1947. Pp. 122.

LUNEBERG, R. K. *Mathematical analysis of binocular vision*. Princeton: Princeton Univ. Press, 1947. Pp. vi+104. \$2.50.

MICHOTTE, A. *La perception de la causalité. Études de Psychologie*, 1946. Vol. VI. Pp. viii+296.

NACHMANSOHN, D., *et al.* *The physico-chemical mechanism of nerve activity*. Annals of the New York Academy of Sciences, Vol. XLVII, Art. 4. Pp. 375-600.

NICOLE, J. E. *Psychopathology. A survey of modern approaches*. (4th Ed.) Baltimore: Williams & Wilkins, 1946. Pp. vii+268. \$4.75.

RADKE, MARIAN J. *The relation of parental authority to children's behavior and attitudes*. Institute of Child Welfare Monograph, No. 22. Minneapolis: Univ. of Minnesota Press, 1946. Pp. x+123.

ROBINSON, F. P. *Effective study*. New York: Harper, 1946. Pp. ix+262.

SADLER, W. *Mental mischief and emotional conflicts*. St. Louis: C. V. Mosby, 1947. Pp. 396.

SANDOW, A., *et al.* *Muscular contraction*. Annals of the New York Academy of Sciences, 1947, Vol. XLVII, Art. 6. Pp. 665-930.

SHERIF, MUZAFER, AND CANTRIL, HADLEY. *The psychology of ego-involvements*. New York: John Wiley, 1947. Pp. v+525.

SNYDER, W. U. (ED.) *Casebook of non-directive counseling*. Boston: Houghton Mifflin, 1947. Pp. viii+339. \$3.00.

SOROKIN, P. A. *Society, culture and personality: their structure and dynamics*. New York: Harper, 1947. Pp. v+742.

THORNDIKE, R. L. (ED.) *Research problems and techniques*. Army Air Forces Aviation Psychology Program, Research Reports. Report No. 3. Washington: U. S. Government Printing Office, 1947. Pp. viii+163.

THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univ. of Chicago Press, 1947. Pp. xix+535.

WOLFF, WERNER. *What is psychology?* New York: Grune & Stratton, 1947. Pp. vii+410.

WOODWORTH, R. S., AND MARQUIS, D. G. *Psychology*. (5th Ed.) New York: Henry Holt, 1947. Pp. iii+677.

WORTIS, S. B., et al. *Physiological and psychological factors in sex behavior*. Annals of the New York Academy of Sciences, 1947, Vol. XLVII, Art. 5. Pp. 603-664.

L'Année Psychologique. Henri Piéron (Ed.). Forty-third and forty-fourth years (1942-1943). Paris: Presses Universitaires de France, 1947. Pp. 856.

Archivio di Scienza della Cerebrazione e dei Psichismi. M. Levi-Bianchini (Ed.). Vol. I-II, 1944-45. Pp. 105-208.

1946 fall testing program in independent schools and supplementary studies. Educational Records Bulletin, No. 47. New York: Educational Records Bureau, 1947. Pp. x+58.

Revista de Psicologia General y Aplicada. Vol. 1, No. 1. Madrid: Instituto Nacional de Psicotecnia, 1946. Pp. 316.

Revista do Centro Psiquiatrico Nacional. Vol. I, No. 1. Rio de Janeiro, Brazil: Imprensa Nacional, 1946. Pp. 161.

Theoria. Ake Petzell (Ed.). Vol. XII, Parts I-II. Lund, Sweden: C. W. K. Gleerup, 1946. Pp. 133.

Psychological Bulletin

THE HISTORY AND PRESENT STATUS OF THE LAW OF EFFECT¹

LEO POSTMAN

Indiana University

Throughout history those concerned with the control of human behavior—parents and educators, businessmen and lawmakers—have acted on the belief that rewards and punishments are powerful tools for the selection and fixation of desirable acts and the elimination of undesirable ones. This commonsense view of the nature of learning has not received undivided support from the professional students of behavior. Almost forty years after the first enunciation of the law of effect as a formal doctrine of learning, the problem of reinforcement is still the subject of heated controversy among the proponents of rival theories of learning. The question has been asked time and again whether this conflict is irreconcilable (47, 142). Perhaps there is some truth in all, or at least some, of the conflicting views, and the main source of difficulty may be the proneness of many theorists to attempt an explanation of *all* learning in terms of their favorite hypotheses. After years of debate, the possibility of a one-principle theory of learning—whether this principle be contiguity, substitution, reward, or expectancy—is still doubtful. This doubt is greatly strengthened by a survey of the history of the law of effect and an evaluation of its present status.

HISTORICAL SOURCES²

Even though the law of effect is invariably associated with the name of Thorndike, who first used this phrase (264), the principles embodied in the law have a long history antedating by many years the modern laboratory studies of learning. The formulation of the law reflected

¹ The writer wishes to express his great indebtedness to Professor Gordon W. Allport who first suggested this review. His unceasing interest and generous advice have been invaluable.

² The law of effect was previously reviewed in this journal by Waters (312). The writer has greatly benefited from this discussion. For the sake of a unitary presentation some of the problems raised by Dr. Waters are discussed again in this paper.

the impact on psychology of three major trends in the history of thought: associationism, hedonism, and evolutionary theory.

Associationism. Ever since Aristotle first wrote of the laws of association of thoughts (similarity, contrast, and contiguity), many philosophers and psychologists have treated the problems of learning, memory and coherent thinking in terms of associations formed among mental elements. Hobbes, Locke, Hartley, Hume, Mill, Wundt, and many more subscribed to the doctrine of association of ideas. Some pinned their faith on the principle of contiguity alone; others introduced similarity as a basic law; some writers, such as Hartley, suggested physiological foundations for the streams of associations. But in spite of divergencies about this or that aspect of the doctrine, they all agreed that mental organization resulted from the joining and weaving together of elements—for a long time the term *idea* predominated—into the content of consciousness. Associationism is, almost by definition, elementaristic and connectionistic. It is elementaristic because it needs conceptual elements which get associated. It is connectionistic because it deals with the principles of connection among these elements. These two basic characteristics of associationistic thought are made explicit in the law of effect.

Hedonism. Associationism is concerned with the laws of connection of mental elements: it has little to say about the role of motives in the acquisition of learned responses. Yet the study of learning cannot proceed long before coming face-to-face with the problem of motivation. In its approach to motivation, modern learning theory has deep roots in the philosophy of hedonism, and the development of the psychology of learning has been characterized by a stubborn defense of hedonistic principles on the one hand and a struggle for the emancipation from hedonism on the other. To the hedonist, pleasure and pain are the governing principles of behavior. The search for pleasure and the avoidance of pain are the mainsprings of conduct, individual and social, and the basis of social interaction and organization. The idea is very old indeed. Plato saw in pleasure and pain important motives of human action; Aristotle called them the basis of the will. Hobbes' interpretation of social life revolved around man's seeking of pleasure and avoidance of pain. The pleasure-pain principle finally found its most systematic expression in the doctrine of utilitarianism (Bentham) which regarded self-interest as a sufficient principle to account for most of individual and social action. The pleasure-pain principle was still very much alive when the law of effect was formulated.

Evolution. The doctrine of evolution served to bring associationism

and hedonism more closely together. The problem of *adaptive* behavior became central. Some responses to environmental stimuli are superior to others and lead to the survival of the fittest. The question at once arises how these superior adaptive responses are *selected* from the multiplicity of responses of which an organism is capable, and then *fixated* and perpetuated. To those who tried to answer this question, hedonism and the pleasure-pain principle provided the principle of selection, and the laws of association the mechanism of fixation.

EARLY FORMULATIONS OF A LAW OF EFFECT

The convergence of associationsim, hedonism, and evolutionary doctrine is exemplified in the formulation of what is in substance a law of effect by Spencer (250) and by Bain (8). Their theories have been reviewed and analyzed in detail by Cason (33) but since they embody salient features of the effect doctrine they will be briefly summarized here. Spencer's (250) basic assumption is that in the course of natural selection there has been established in the various species a correlation between the pleasant and the beneficial, and a similar correlation between the unpleasant and the injurious.³ That which is pleasant is maintained and repeated and proves beneficial to the biological organism. That which is painful is abandoned and the organism is protected from injury. Spencer then described a physiological mechanism for the selection of useful pleasant acts and the elimination of harmful unpleasant acts. Organisms respond to environmental stimuli in a highly variable, essentially *random* fashion with diffuse discharges of neural energy. Among these random responses there will be sooner or later one (accidental) response which is successful (for example, produces food for the animal). After success will immediately come pleasurable sensations with a large discharge of nervous energy toward the organs engaged in the successful act, e.g., in eating. This heightened discharge of nervous energy will render the successful channels of muscular action more permeable. On recurrence of the circumstances the neural discharge will no longer be diffuse but will be channeled into the successful movements. With every repetition the successful channels will be made more and more permeable, until stable nervous connections have been organized.

³ This assumption is echoed in James' *Principles* where we read: "If pleasure and pains have no efficacy one does not see why the most noxious acts, such as burning might not give thrills of delight, and the most necessary ones, such as breathing cause agony" (132, p. 143). Another exponent of a pleasure-pain theory of learning was J. Mark Baldwin (9). See also Pyle (215).

Bain (8) formulated a similar theory: spontaneously begun movements which are accidentally successful cause pleasure, and with the pleasure there is an increase in vital energy. A few repetitions of this coincidence of pleasure and movement lead to a neural connection so that pleasure or the idea of the pleasure will evoke the successful movement at once. In the case of pain, the sequence is the reverse, leading to a decrease of vital activity and a blocking of movement.

The theories of Spencer and Bain have been summarized in some detail here because certain salient features of the pleasure-pain theory have proved extremely persistent in later developments:

1. The organism is said to respond to the problem situation by *random* movements and spontaneous discharges of the muscles. Later analyses of trial-and-error responses have been attacked because they seemed to imply that an animal's initial responses in a problem situation are random (148). This would be an extremely hazardous assumption to make, for in a problem situation an animal will rarely respond completely at random; even his trial-and-error responses always have a certain directionality though they may at first appear random from the observer's viewpoint. Thoughtful analysts of trial-and-error learning have recognized this fact and pointed out that the assumption of randomness is not necessary for a trial-and-error theory of learning (240).

2. The causal efficacy of pleasure and pain in fixating successful responses and eliminating unsuccessful ones is explicitly asserted. Pleasure and pain are purely psychic concepts and mechanistically inclined thinkers have been reluctant to endow them with such power over muscular responses (29, 33, 112). Hence the need to invoke physiological correlates of pleasure and pain to reduce their action to mechanistic principles. This need was felt not only by Spencer and Bain but also by later writers continuing the tradition of a pleasure-pain theory of learning (252, 264, 297). Most of these physiological explanations have been highly speculative, as are certainly those advanced by Spencer and Bain. There is, for example, no evidence to support the contention that pleasure is accompanied by heightened neural activity and pain by lessened activity. This may sometimes be the case but the opposite may be equally true, especially in the case of intense pain. The trouble was that a pleasure-pain theory was really a theory of psychic causes leading to physical effects. Physiological explanations have been largely ad hoc and invoked for the sake of consistency with basic axioms about the mechanistic nature of the learning process.⁴

3. The third feature which the pleasure-pain view bequeathed to modern learning theory is its insistence that a repetition of a successful response will

⁴ The vexatious issue of hedonism was sidestepped by Hobhouse (110) who described effect in terms of confirmation and inhibition. Similarly Holmes (114) speculated that the congruity or incongruity of an act with the activity in progress constituted its effect. Such formulations, however, only postpone the consideration of the role played by satisfaction and annoyance. As Stephens (257, 258) has pointed out, if an organism unexpectedly stumbles on a valuable outcome, the failure of the response to confirm the expectancy will not prevent the response from recurring. Similarly, a congruous expected response may be eliminated if it leads to unpleasant consequences.

strengthen it. There are many cases where this is true but probably equally many where the opposite is equally true, where continued repetition of a response will weaken that response (59).

4. Let us finally note the central emphasis which a pleasure-pain theory puts on the motor (movement) aspect of the connection between situation and response. It is the movement which is strengthened by pleasure and increased neural activity. Later formulations of the law of effect, especially in conjunction with conditioning principles (127) have placed equal emphasis on the strengthening of afferent-efferent connections by reinforcement. Critics have insisted that movements as such are rarely learned and used by animals only as means to an end. It is probably safe to say that changes in the *stimulus as the animal perceives it and interprets it* (3, 29, 30) are just as essential as the fixation of successful movements.

THE NEW ASSOCIATIONISM: THORNDIKE'S LAWS OF LEARNING

The combination of a modified pleasure-pain (success-failure) philosophy and connectionism found its most systematic and challenging expression in the work of Thorndike. Even though he later disclaimed hedonism (273), Thorndike's thinking has reflected that of the earlier exponents of the pleasure-pain principle as well as the associationist doctrine. His central importance derives from the fact that he embarked on a monumental series of experiments, extending over half a century, to obtain empirical verification of his laws of learning. Partly on the basis of his study of the problem-solving behavior of animals, Thorndike formulated two basic laws of learning: *the law of exercise* and *the law of effect*.

Other things being equal, the *law of exercise* makes learning a function of the number of repetitions of the stimulus-response connections.

The law of effect was stated as follows: "Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that when it recurs, they will be more likely to recur; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connection with the situation weakened so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort the greater the strengthening or weakening of the bond" (264, p. 244). More succinctly expressed, the law as originally stated asserts that success stamps in and failure stamps out.

In the light of careful experimental analysis Thorndike later abandoned the law of exercise, or at least he relegated this law to a very minor position, and for him the law of effect became altogether central to the learning process (267, 269). At various times Thorndike's statements of the law of effect have been accompanied by an exposition of its probable physiological correlates. The physiological locus of rein-

forcement by effect is put squarely in the neurons (e.g. 264, 267, 269, 270, 273). Satisfying and annoying processes are "related to the maintenance and hindrance of the life processes of the neurons" in whose changing synaptic conduction learning consists (264, p. 24). In other words, pleasure or satisfaction facilitates synaptic conduction, renders synapses more permeable, whereas discomfort or pain blocks synaptic conduction.

The subsequent history of the law of effect has shown two major foci of interest.

1. The first major focus of interest has been Thorndike's law itself. Ever since its first formulation it has been subjected to a long series of attacks, attacks which were vigorously countered by Thorndike and his associates both on the level of theoretical discussion and in the experimental laboratory. In the course of these discussions Thorndike himself has shifted his theoretical position to some extent and reformulated the law. Concurrently many of the important parameters of reinforcement (such as amount of reward, frequency of reward, delayed reward, spread of effect, etc.) were investigated.

2. A second focus of interest has developed around the integration of the law of effect with the facts and theories of conditioning. In Hull's systematic theory of learning (127), the law of effect occupies a central position. On the other hand, the law of effect has been singled out for attack by those who are opposed to a conditioning-reinforcement type of learning theory (e.g., Tolman and his associates: 289, 294, 315).

The law of effect has thus become an issue around which systematic differences in behavior theory have been crystallized. Defense and condemnation of the law of effect have almost become symbolic of widely divergent approaches to the problems of behavior analysis and learning theory. The discussion which follows will be centered around these two foci of interest and will also attempt to cover a representative selection of studies dealing with the parameters of reinforcement.

THE CLASSICAL OBJECTIONS TO RETROACTION

Let us turn first to a consideration of the classical objections to the law of effect. In the language of the law of effect, the consequence of a response strengthens the response, i.e., the effect "works back" on the connection which it follows. Such a formulation involves the assumption of retroaction. Critics asked, how can an effect work upon a response which is already passed (29, 33, 46, 212, 219)? The logical difficulty is obvious, and Thorndike has not been inclined to minimize it. He has suggested that the physiological equivalent of the connection does not vanish instantaneously but is still present when the consequence (reward or punishment, knowledge of results) occurs (269, p. 481). The original statement concerning backward action should then be modified

to read that one type of after-effect (satisfier or annoyer) acts on other after-effects of a connection (the persisting physiological equivalent of the original connection).

The same logic for bridging the temporal gap between connection and consequence and avoiding the cul-de-sac of retroaction underlies Hull's use of the concept of *stimulus-trace*. Upon cessation of a stimulus (exteroceptive or proprioceptive) the afferent impulse continues its activity for a finite short period, and a short-range temporal integration is achieved. Hull writes, "This perseverative stimulus-trace is biologically important because it brings the effector organ en rapport not only with environmental events which are occurring at the time but with events which occurred in the recent past, a matter frequently critical for survival" (127, p. 385). Not only may the physiological correlate of a stimulus or response persevere in time but symbolic processes may help the human learner and, to a lesser extent, the animal learner to bridge the temporal gap between the occurrence of the connection and the action of the satisfier (192).

There have been other attempts to slip from under the horns of the dilemma of retroaction. Accepting the apparent fact of backward action at its face value, retroflex circuits in the brain (297), neural irradiation (96), changes in the electrical resistance of neural connections (251) and their readiness to conduct (252, 267), the rearousal of just-active pathways (44)—all these were suggested as possible mechanisms mediating the action of satisfiers and annoyers. These physiological explanations can be called neither right nor wrong. They are clearly speculative and many steps removed from the level at which experimental verification is at present possible.

The most radical solution of the problem has been to deny retroaction altogether and to affirm that in spite of the appearances to the contrary effect always operates in a forward direction. What is modified is not the connection which is passed and done with but the stimulus on the occasion of its next appearance. "The burned child shuns the fire not because pain did anything to his movements, but because, since that pain, the stimulus has changed; it is now flame plus fear, no longer flame plus curiosity" (112, p. 218f.). Similarly, Carr ascribed the fixation and elimination of responses primarily to the sensory consequences of an act rather than to the strengthening of S-R connections. "These consequences do not influence the portion of the act that preceded them . . . they do affect the subsequent functioning of the act" (29, p. 96). Whether or not it is possible to cut the Gordian knot of retroaction by insisting that effects exert their influence in a forward direc-

tion, criticisms such as those of Hollingworth and Carr were important because they shifted the emphasis from the *response* to the *stimulus*.

In trying to evaluate the arguments for and against retroaction, one is struck by the fact that the battle has been fought largely on the plane of logic (312). There can be no doubt that rewards and punishments following responses modify these responses on the occasion of their recurrence. Not enough is known about the basic mechanisms of learning to decide whether consequences of a response can or cannot act back on a connection or whether perforce they must exert their effect in a forward direction. Operationally speaking, the consequences of a response always act in a forward direction since they can only be tested on the occasion of a future occurrence of the stimulus situation. Physiologically speaking, when all is said and done we had best admit our ignorance of the action of cortical neurones. Perhaps part of the confusion has stemmed from a premature reification and physiologizing of the terms involved. The words *connection* and *effect* are logical constructs, the words *stimulus* and *response* are generic terms denoting classes of events with certain properties in common (237). It is only too tempting to forget the limitations of these concepts and to manipulate abstractions as if they were clearly delineated entities with equally unequivocal physiological counterparts. For analytical purposes we need such concepts as *stimulus*, *response*, and *connection*, but are we not over-optimistic when we assume that the nervous system makes its division along the same lines? If misplaced physiological concreteness is avoided one may escape logical difficulties such as the dilemma of retroaction.

WHAT IS THE NATURE OF SATISFIERS?

Whereas some critics were most concerned with the mechanisms mediating effect, others focussed their attention on the nature of the satisfiers and annoyers to which reference is made in Thorndike's law. Although Spencer and Bain, in whose tradition Thorndike continued, frankly invoked pleasure and pain as agents responsible for the fixation and elimination of responses, Thorndike's law has been a law of *effect*, not *affect* (113). He carefully defined satisfiers and annoyers in terms independent of subjective experience and report. "By a satisfying state of affairs is meant one which the animal does nothing to avoid, often doing such things as to attain and preserve it. By a discomforting state of affairs is meant one which the animal avoids and abandons" (264, p. 245). Although admittedly free of hedonism, such a definition of satisfiers and annoyers has faced another serious difficulty: the danger of circularity. The critic may easily reword the definition to read: "The

animal does what it does because it does it, and it does not do what it does not do because it does not do it." This *reductio ad absurdum* is probably not entirely fair but it points up the danger of the definition in the absence of an *independent* determination of the nature of satisfiers and annoyers. The satisfying or annoying nature of a state of affairs can usually be determined fully only in the course of a learning experiment and cannot then be invoked as a causal condition of learning without circularity (108). In their experimental work Thorndike and his associates have made no significant attempts to establish the satisfying or annoying nature of their rewards and punishments independently of the learning experiment. The tacit assumption has been made throughout that the words "Right" and "Wrong" and/or small money gains and losses are, indeed, satisfying and annoying states of affairs. As Tolman and his co-workers have argued on the basis of strong experimental evidence, such events (including shocks) can equally well be conceived as signals or "emphasizers," providing the subjects with information as to the correct response (294). We shall return shortly to the problem of cognitive (informative) vs. satisfying and annoying after-effects of a connection.

Thorndike's definition of satisfiers and annoyers has been defended against the stigma of circularity by emphasizing its empirical adequacy (171). "The primary fact is that there is a state of affairs, whatever its psychological classification, which happens as a result of, or at least after, an act. The state of affairs may be symbolic . . . or it may be primarily sensory . . . or it may be complexly perceptual" (171, pp. 576-577). This statement is fully in line with Thorndike's own reformulation of the nature of a satisfying state of affairs. In his more recent writings Thorndike has conceptualized the action of satisfiers in terms of the *O.K. reaction* (273, 277), a fundamental confirming reaction brought to bear by the organism on connections between situations and acts. He describes the O.K. reaction as the "unknown reaction of neurones which is aroused by the satisfier and which strengthens connections on which it impinges" (273). This reaction is independent of sensory pleasure. It is "far from logical." It strengthens "connections which are wrong, irrelevant, and useless." It is said to bear little relation to the intensity of the satisfier. Neither specific motivation nor effects relevant to a specific motivation are any longer assumed.

Stripped of virtually all defining properties and qualifications, the law does indeed have a very wide range of applicability but only at the expense of vagueness. The sum and substance of the argument now is that something happens in the organism (nervous system) after an act

is performed. The fact that something happens influences further action. This something is, however, so little defined that it has almost no predictive efficiency. The O.K. reaction has no measurable properties, the conditions for its occurrence are so general as to embrace almost every conceivable situation. Hence the operation of the O.K. reaction can be inferred only *ex post facto*, after learning has taken place. But here we are impaled again on the horns of the dilemma of circularity.

Perhaps because the reformulated empirical law of effect had a degree of generality which made it an unsatisfactory tool for the prediction of specific facts of learning, several auxiliary principles were gradually introduced in the connectionist framework: the principle of *belongingness*, according to which S-R correlations which "hang together" are learned more easily than connections lacking this quality; the principle of *impressiveness* which states that vivid stimuli are favored in learning; the principle of *polarity* according to which stimulus-response sequences function most readily in the order in which they were practiced; the principle of *identifiability* stating that the most easily identifiable connection is most easily learned; and finally the principle of *availability* which has it that the more available a response the more easy it is to connect it to a stimulus (227, 269). It is easy to see that most of these principles are closely akin to the concepts of Gestalt psychology. According to Brown and Feder (18), "these conditioning factors which for Thorndike are so many *ad hoc* hypotheses, are logically prior to his other laws." These writers claim that Thorndike's theory of learning could be successfully rewritten in terms of Gestalt psychology. Whether or not this assertion is true,⁵ the use of these auxiliary principles serves to point up the inadequacy of the O.K. reaction as a universal principle of learning.

It is very probable that the modifications in Thorndike's description of the process of reinforcement were largely influenced by the results of his own experimental labors over the years. Most of his experiments were concerned with rote verbal learning and the acquisition of simple skills (269), and it is the results of these experiments which were interpreted in terms of the confirming reaction and such auxiliary concepts as belonging and identifiability. Other advocates of the law of effect, especially those whose experimental interests have been more in the fields of conditioning and discrimination learning, have tended to insist

⁵ Certainly Thorndike would not agree. He professes his "inability to understand" Gestalt psychology. He feels suspicious of the concepts of Gestalt psychology "because for a time they were offered as a substitute for psycho-vitalism" (267, p. 125). He concludes that a connectionist theory is "far simpler and more in accord with what the neurons are and can do" (p. 131).

on a version of the law in which the term *effect* retained much more of its original meaning and had direct reference to the primary and secondary needs of the organism. Let us consider, for example, a recent formal statement by Hull of his *law of primary reinforcement* which, in his own words, is distinctly related to the law of effect. The law reads as follows:

Whenever an effector activity occurs in temporal contiguity with the afferent impulse, or the perseverative trace of such an impulse, resulting from the impact of a stimulus energy upon a receptor, and this conjunction is closely associated in time with the diminution in the receptor discharge characteristic of a need, there will result an increment to the tendency for that stimulus on subsequent occasions to evoke that reaction (127, p. 80).

Reduction of a need is a critical factor in the reinforcement process, and by *need* Hull means a condition in which any of the commodities necessary for the survival of the organism are lacking or deviate seriously from the optimum (127, p. 17). Disturbances in the "optimal conditions of air, water, food, temperature, intactness of bodily tissue, and so forth" give rise to needs and activities of the organism for the search of survival. It is the highly important adaptive role of the learning process to reinforce those movements which will lead to need reduction and thus help the animal to survive. Hull recognizes, of course, that much learning takes place in the absence of immediate reduction of biological needs. To account for such learning, the principle of secondary reinforcement is invoked. "The power of reinforcement may be transmitted to any stimulus situation by the consistent and repeated association of such stimulus situation with the primary reinforcement which is characteristic of need reduction" (127, p. 97). Thus *all* reinforcement depends directly or indirectly on the reduction of "viscerogenic" needs. Since reinforcement is a basic condition of learning, it follows that learning is primarily an adaptive process in the interest of species survival. Biological needs provide the foundation on which the complex structure of a lifetime of learning is built.

Hull's formulation of the process of reinforcement has been echoed by many other investigators in the field of learning. In a long series of experiments (some of which will be discussed in detail below) O. H. Mowrer, for example, has attempted to produce experimental evidence and theoretical argument for a version of the law of effect which is closely parallel to that of Hull in its emphasis on tension reduction. A typical statement by Mowrer claims that

... learning is dependent not upon the mere association, or temporal contiguity, of stimuli (or responses) but rather upon the occurrence of a state of affairs which has been variously designated as goal-attainment, problem-solution,

pleasure, success, gratification, adjustment, reestablishment of equilibrium, motivation-reduction, consummation, reward (190).

Elsewhere Mowrer makes it unequivocally clear what he means by satisfaction. He writes:

Most writers agree that satisfaction ought to be equivalent to pleasure and dissatisfaction equivalent to pain. But, then, what do we mean by "pleasure"? One view is that we can equate pleasure to drive or tension-reduction and pain to drive or drive-increase. . . . Satisfaction, pleasure, and drive-reduction are strictly equivalent (185).⁶

Mowrer's interpretation of effect or a satisfying state of affairs is frankly and emphatically hedonistic.⁷ Ultimately all behavior modification is mediated by pleasure (tension-reduction). The statement has a very familiar ring. The history of learning theory has completed a cycle, and the views of Spencer, Bain, and Baldwin are with us again though transplanted into the conditioning laboratory and accompanied by great refinement and sophistication in experimental techniques and procedures. The historian of learning will be especially interested to note that while Thorndike, the original advocate of the law of effect, has frequently dissociated himself from a hedonistic interpretation, there has emerged a group of *neo-hedonists* whose interpretation of the law is much closer to its historic roots. They were willing to assume the responsibility of hedonism in order to make the law more useful in the design and interpretation of experiments.

To many psychologists hedonism today is no more acceptable than it was at the turn of the century. A vigorous attack against the preservation of a hedonistic doctrine in the law of effect has been leveled by Allport, who has been especially concerned with the implications of a law of effect for the growth and development of personality. A strict adherence to a hedonistic law of effect would imply that "personality grows through the production of tentative trial-and-error acts, some of which are selected by the Great God Pleasure for establishment, and some by the Great God Pain for banishment" (2, p. 154). Allport believes that the main fallacy of hedonism is the confusion of the *by-product* of a complex process with the process itself. He would agree that pleasure (tension-reduction) at times is an indicator of the successful accomplishment of an act (4) but emphatically denies that pleasure

⁶ Similar emphasis on tension-reduction as the basic process of reinforcement is found in the writings of Meunzinger (199) in connection with his work on the effect of electric shock on learning.

⁷ There is an important way in which such a view differs from that of the classical hedonists. For the hedonists pleasure was a motive, whereas for the modern learning theorists it is tension, or discomfort, which is the motive while pleasure is the result of tension-reduction which leads to learning (192).

is the main causative factor in the fixation of responses. He also takes issue with the law of effect because it perforce implies a segmental approach to the learner, ascribing learning to the satisfaction of this drive or that drive with the main emphasis on the reinforcement of *responses* (3). For Allport it is the person who learns, is rewarded, and utilizes his success in his future adjustment. Allport thus pitted an ego-oriented psychology of learning against the neo-hedonism of the law of effect. To this controversy we shall return presently.

In attempting to evaluate the controversy which has raged around the definition of satisfiers one is struck by the key importance of the hedonistic issue. Certainly hedonism is an immediate ancestor of the law, and now that the principle of effect has reached an uneasy maturity it is clear that it cannot deny its origin without sacrificing much of its vigor. When the law is stripped of hedonistic implications, when effect is not identified with tension-reduction or pleasure (as by Thorndike), the law of effect can do no more than claim that the state of affairs resulting from a response in some way influences future responses. Such a statement is a truism and hardly lends itself to the rigorous deduction of hypotheses and experimental tests. If a neo-hedonistic position is frankly assumed (as, e.g., by Mowrer) the law becomes an important tool for research, provided "satisfaction" is independently defined and not merely inferred from the fact that learning has occurred. The hedonistic position can be defended for the range of events for which it is experimentally demonstrable. It is probably fair to say that up to the present this range is fairly narrow and largely in the domain of animal learning.

The main source of difficulty has been the universality claimed for tension-reduction as a principle of reinforcement. Hull and Mowrer, on the one hand, and Allport, on the other, probably would agree that reduction of the food need is a critical factor in a rat's learning the true path through a maze. With the aid of the principle of secondary (and higher order) reinforcement, tension-reduction can then be made the ultimate condition of all learning. Such a generalization Allport rejects. Crucial experimental tests to arbitrate between these two interpretations of more complex forms of learning still seem to be lacking. For the time being, therefore, acceptance or rejection of the hedonistic postulate must remain a matter of philosophical preference.

THE NATURE OF PUNISHMENT

Thorndike's View of the Effect of Punishment

The preceding discussion has dealt primarily with the nature and action of satisfaction (reward). As originally stated, the law of effect

was a two-pronged statement, putting the stamping-out powers of punishment on a par with the stamping-in effects of reward. In the light of experimental results Thorndike no longer considers punishment as an effective agent for the elimination of wrong responses. Indeed, again as a result of experimental evidence, a punished connection is now held to be strengthened more by sheer occurrence than it is weakened by punishment. Punishment is not the dynamic opposite of reward; it is not even capable of overcoming the effect of sheer occurrence (269, 279). This view is, of course, an implicit resurrection of the law of exercise since sheer occurrence (frequency) is now credited with a certain degree of effectiveness in fixating responses, an effectiveness, moreover, which surpasses that of punishment.

An impressive number of experimental investigations have been conducted by Thorndike and his associates to demonstrate the ineffectiveness of punishment (162, 163, 165, 166, 223, 268, 269, 276, 299). In most of these experiments the punishment consisted of the announcement *Wrong* made by the experimenter, accompanied sometimes by a small fine or an electric shock. The effect of the punishment was assessed in terms of the number of repetitions of the punished response as compared to the number of repetitions that would be expected on the basis of chance, i.e., if the subject were guessing at random. The results of the experiments led Thorndike to conclude that punishment, instead of weakening or "stamping-out" the wrong response, may have a variety of effects depending on the specific nature of the annoyer and the propensities of the organism. The important point is that what an animal is led directly to do by an annoyer may or may not make the repetition of the punished response less likely (269). If punishment does lead to the elimination of a response, its action is *indirect*: it leads to variability of behavior, thus increasing the opportunities for the occurrence of the correct response which is then reinforced by the direct action of the satisfier (OK reaction).

Thorndike's revised view of punishment startled the psychological public. Not only did it contradict the belief in the practical value of punishment which had become almost an axiom in our social life, it was also clearly at variance with a considerable body of other experimental data. We shall now consider (1) the criticisms leveled against Thorndike's work on the effect of punishment, and (2) other theoretical and experimental developments bearing on the role of punishment in learning.

Criticisms of Thorndike's View of The Effect of Punishment

One serious criticism of Thorndike's data was statistical in nature.

The effect of punishment was measured in terms of the deviation of the obtained number of repetitions from the number of repetitions to be expected by chance. Thus, if there were five possible responses to an item, Thorndike would consider the chance expectation of each response to be .20, and the obtained percentages were expressed as deviations from .20. Thorndike was thus applying the principle of indifference which considers alternative events equally likely in the absence of information to the contrary. The trouble is that there seems to be some information to the contrary, viz., that the "natural choice frequency is usually not a chance one" (123). Whether or not Thorndike's correction for chance is adequate is an empirical question. It is necessary to compare the extent of actual repetition without after-effect with the extent of repetition following announcement of *Wrong*. Experimenters who used an empirical rather than an a priori baseline for the evaluation of the effects of *Wrong* find that punishment does have a weakening influence which is commensurate with the strengthening influence of reward (254, 255, 283). Thorndike and his associates fail to find such an effect even when favoritism of response is taken into account (162, 166).

Failing to find a significant weakening of connections following the announcement of *Wrong*, Thorndike denied the efficacy of punishment in general.⁸ Yet even within the restricted framework of a typical Thorndike experiment, and with as highly specific a "punishment" as the announcement of *Wrong*, the results vary considerably with the parameters of the experimental situation. The medium by which the punishment is conveyed is important: it may make a difference whether *Wrong* is announced by spoken word or by, say, a signal light. The effect of the medium, the sheer "something happening" after a response may be powerful enough to obscure the weakening influence of punishment. Such, at least, was the conclusion of Stephens who argued that whenever punishment strengthens a bond, such strengthening may be due to the physical medium, whereas reward and punishment have directly opposite effects when measured from the baseline of informationless after-effects (a flash of light, a nonsense syllable) (255).⁹ Jones, (139) on

⁸ While the majority of experiments were done with human subjects, Thorndike illustrated the superiority of reward also with animal subjects (268).

⁹ Thorndike and his associates also found that informationless after-effects such as a click led to more frequent repetitions than the announcement of *Wrong*. They were unwilling to conclude, however, that punishment had weakened a response and argued that a "neutral" after-effect was ambiguous enough to lead to self-administered reward since the subjects were free to interpret the neutral signal as a reward (165, 166). Such an argument comes dangerously close to question-begging. It indicates an unwillingness even to consider the possibility that punishment may have a weakening effect on a connection.

the other hand, found that the medium by which punishment is conveyed makes little difference and that the elimination of errors depends on the *position* of the punished response in the pattern to be learned.

Whatever the effect of the medium by which the punishment is conveyed, the initial strength of the association will be an important determinant of the effect. A strong association can be considerably weakened, a weak association which is close to zero cannot be interfered with much further. Stephens was able to show experimentally that *strong* connections (responses which appeared early and were persistent) are markedly and dependably weakened by the announcement of *Wrong* whereas initially weak ones are not. Correspondingly, *Right* has little, if any, influence on initially strong associations but considerably aids in making weak bonds strong (254, 256). Tilton (284, 285) also emphasizes the importance of taking the initial strength of a response into account.

There is another feature of the experimental situation which Thorndike fails to take into account but which may be, at least in part, responsible for his failure to find a direct effect of punishment. In most of the experiments, there was only *one* right response and several wrong alternatives. Thus it was the subject's task to learn positively only one item, but to eliminate several items. By virtue of its uniqueness, the right response is "figural" (has, as it were, high visibility for the subject), whereas a wrong response is one in a long homogeneous series of wrong responses (213). These considerations receive experimental support from the work of Dand (45) who equated the number of right and wrong alternatives and presently found an announcement of *Wrong* to have a definite weakening effect. Lorge (162) reports that the potency of a punishment is less the higher the initial probability of obtaining a right response (though not reliably so). It is safe to say that the pattern and sequence of *Rights* and *Wrongs* are at least partial determinants of the fixation and elimination of responses.

The complex ways in which the effects of *Wrong* vary with the parameters of the experimental situation highlight the need for caution in generalizing about the effects of punishment. Even when the punishment consists of a simple announcement of *Wrong*, the experimental evidence is still inconclusive. An announcement of *Wrong* is only a very special type of punishment situation. Indeed, one may question the appropriateness of the term *punishment* or *annoyer* for the description of an effect which is primarily informative in nature. In most of Thorndike's situations, the connections to be learned are purely arbitrary and it is easily obvious to the subject that being right or wrong cannot possi-

bly reflect on his intelligence. He may be neither gratified by being right nor annoyed by being wrong. *Wrong* may merely serve as a signal to substitute one arbitrary guess for another. We must reiterate the need to define satisfaction and annoyance independently before invoking them as determinants of the learning process.

Other Studies of the Role of Punishment

When we turn to punishments other than an announcement of *Wrong* or slight monetary losses, the uncertainty of Thorndike's generalization becomes even more apparent. An impressive array of experimental evidence can be marshalled in support of the proposition that punishment is an effective condition of learning. We shall first review the empirical findings and shall then consider the main theoretical issues concerning the nature of punishment and the mechanisms of its action.

Society has devised many and varied punishments for those who offend against law and convention but in the psychological laboratory punishment has come to be almost synonymous with electric shock. Shock is easily administered and it is always sure to have a physiological impact though its psychological effects are highly variable (296). Thus most of the data on punishment come from studies on electric shock.

A review of the literature on punishment raises the following problems:

1. Does punishment have any significant effect on learning?
2. Is the effect of punishment generalized to the entire learning situation or does it lead only to the elimination of specific punished responses?
3. What is the mechanism by which punishment exerts its effects? Does punishment weaken stimulus-response bonds? Does it affect the general level at which the organism functions? Or is its effect indirect, leading to increased variability of behavior and ultimately to differential reinforcement by reward?
4. Does the main effectiveness of punishment reside in the *information* about errors with which it provides the learner or does it have a more immediate influence on associative strength?

Does punishment have any significant effect on learning? We have already seen that Thorndike answers this question in the negative but that the validity of this conclusion from his data is doubtful. The answer to this question is, trite as it may sound, yes and no. It is rather unfortunate that Thorndike and his associates have tended to put the issue on an all-or-none basis. The question should not be, "does punishment have an effect on learning or does it not?" but rather, "under what conditions is it effective and under what conditions does it fail to show results?"

Against the blanket assertion that punishment is not instrumental

in the elimination of wrong responses it is possible to cite a long list of papers covering more than half a century of experimental work which report that punishment is an effective condition of learning.

Without attempting to review these experiments in detail, we shall summarize the main findings.

1. A large number of different organisms have been spurred on to better learning by punishment:¹⁰ earthworms (325), cockroaches (261, 300), dancing mice (324), rats (e.g., 20, 54, 56, 57, 111, 116, 155, 193, 196, 199, 200, 201, 202, 204, 206, 301, 307, 322), chicks (36), cats (53), and, last but not least, human beings (10, 12, 13, 19, 24, 25, 26, 27, 41, 55, 79, 80, 87, 137, 138, 172, 173, 217, 218, 302).

2. Punishment, usually by electric shock, has improved learning in a variety of problem situations: discrimination boxes, mazes, reaction-time procedures, serial learning, mirror tracing, multiple choice situations.

3. Punishment has been found to affect different measures of learning. In many studies, in which responses are scored as either right or wrong, punishment has led to a decrease in the number of errors (10, 13, 24, 25, 26, 27, 36, 41, 54, 55, 56, 57, 79, 80, 87, 111, 116, 137, 138, 172, 173, 193, 196, 199, 200, 201, 202, 204, 206, 218, 301, 302, 307, 326). When speed of learning is measured in terms of the number of trials required to reach criterion, we find that punishment often results in a smaller number of trials (19, 24, 25, 26, 27, 36, 41, 54, 55, 111, 137, 155, 171, 193, 196, 199, 201, 202, 204, 206). Some investigators report the fact that punishment reduces the time required for learning (24, 25, 41, 55, 79, 80, 87, 206, 302) but time has proved to be a difficult measure to interpret since punishment typically leads to a cautious hesitant attitude on the part of the learner so that a decrease in the number of trials and errors may be accompanied by an increase in time per trial (10, 24, 25, 26, 172, 173, 206, 218).

4. The effectiveness of punishment may also be demonstrated by retention tests. There are experimental reports suggesting that (1) punished groups remember the learned task better (27, 41, 56, 79, 80, 81, 206, 301), (2) punished groups are less subject to retroactive inhibition (27). Different measures of retention show the effect of punishment to unequal degrees, however (41, 81, 301).

On the basis of the experimental evidence it is thus possible to assert at least that punishment works in some situations some of the time. But note the qualifier *some*. None of the papers cited would justify the conclusion that punishment works in all situations all of the time. On the contrary, the experimental evidence emphasizes the extent to which the *effectiveness or non-effectiveness of punishment depends on the parameters of the experimental situation*.

1. The intensity of the punishment is important. As Yerkes and Dodson (326) pointed out more than forty years ago, and many investigators since have confirmed, there is in a given learning situation an *optimal* intensity of punishment. The relationship between intensity of punishment and learning efficiency is not linear over the total range of intensities. Once the optimal intensity has

¹⁰ Conditioning studies employing a punishing UCS are not included here.

been exceeded, efficiency of learning may decrease (36, 53, 87, 302, 326, 328). There is moreover, no one optimal intensity for any given organism. The optimal intensity will vary significantly with (1) the difficulty of the task and (2) the stage of learning at which punishment is applied.

2. There has been rather general agreement that relatively severe punishment (intensive shock) is most effective in the learning of simple habits such as black-white discrimination and simple maze patterns, and that relatively mild punishment is optimal in the case of difficult tasks such as more complex types of discrimination and exacting maze patterns (12, 26, 36, 53, 326, 328).

3. The effects of shock may become disruptive (12, 79, 80, 87, 294, 303), especially if applied early in the learning process (25, 206, 218). Shock may also lead to increased variability of responses as compared with milder forms of punishment (260).

4. As to the stage of learning at which punishment is most profitably introduced, we have already referred to Stephens' general conclusion that the greatest effect is achieved with strong connections (254, 256). This conclusion is borne out by the findings of Valentine (301) and by a systematic investigation of Bunch (25) who got better results with a few shocks late in the learning than with many more administered earlier in the process. The results of his experiments led Bunch to believe that for a given task there is an *optimal combination of position in learning and number of trials with shock*. This generalization clearly reflects the dependence of the effect of punishment on the experimental parameters.

Any generalization about punishment based on a single experiment, or even a series of experiments, thus becomes extremely hazardous. For a given organism, there are combinations of conditions under which punishment is effective. If the experimenter works outside this range of combinations, punishment fails to yield significant results (Thorndike and co-workers). If he works with several combinations of conditions, he may easily find that punishment is effective in one case but not in the other and certainly that it varies in effectiveness (12, 26, 87). If he works with only one set of parameters, he may be tempted to over-generalize. Such seems to have been the case in Thorndike's experiments on the effect of *Wrong* since most of the connections subjected to punishment were probably rather weak (254), and the punishment hardly very intensive in nature. Such lack of consistency in results need not lead to a counsel of despair regarding the possibility of generalization about punishment, but merely to an insistence that any such generalization should be based on functional relations covering a wide range of parameters and not on isolated, restricted experiments.

A special problem of control needs to be mentioned here. In an attempt to hold as many of the experimental conditions as possible "constant," the experimenter may decide on a certain type and intensity of punishment and then adhere to it throughout the course of the experiment. He would then overlook the fact that the effectiveness of a punishment changes as a function of the number of times that it has

been administered. In the case of shock, there are clear-cut experimental data to substantiate this point. Investigating electric shock as a motivating stimulus in conditioning experiments, Kellogg (143) found that in order to maintain a flexion reflex of *fixed extent* it is necessary to administer relatively large voltages at the beginning of the series and to reduce the stimulus intensity toward the end of the series. He also found that the intensity required varied with the distribution of practice. In addition, Kellogg reported wide individual differences among his subjects (dogs) in sensitivity and emotional reaction to the same physical shock. He concluded:

It seems reasonable to infer that in learning experiments where the physical properties of an electrical stimulus are constant, the effects produced in the same subject at different times must differ greatly. The attempt to maintain "constant motivation" by this method probably defeats its own purpose (143, p. 95).

Considerable individual differences in reactions to shock have also been reported for human subjects (296). Too few experimenters have heeded Kellogg's admonition to consider the effectiveness of a shock at a given point in time rather than its physical intensity.

Is the effect of punishment general or specific? In the framework of connectionism, it is always a *specific* bond which is strengthened or weakened. Thus Thorndike and his co-workers have always concerned themselves with the history of *specific* rewarded and punished responses, finding punishment ineffective. Other investigators have found, however, that punishment may have a much more general effect: there may be a general improvement in performance, with unpunished responses benefiting along with the specifically punished ones (10, 12). Even if the punishment itself is made non-specific ("non-informative shock"), and not administered as a consequence of any particular response, there is still a facilitating effect (11, 12, 19, 25, 79, 80). The work of Muenzinger is of especial importance here. In a series of carefully controlled experiments he was able to show that (1) shock administered to rats *after* the choice-point in a maze facilitates learning and is more effective than hunger-motivation alone (193, 196); (2) shock *before* the choice-point decreases the efficiency of learning¹¹ (204) as does shock at the moment of choice (71); and (3) if shock is administered throughout the maze, its effect is only slightly less than with punishment *after* the choice point (199). These findings apply to shock for *right* responses as well as shock for *wrong* responses. When both right and wrong responses are shocked, there is no summation of effects though learning efficiency is greater than if either right or wrong choices alone are followed by shock (204).

¹¹ This retardation is probably due to the persistence of position habits which become fixated (70, 101).

The question asked at the beginning of this section, whether the effect of punishment is general or specific, turns out to be an unwise one. Clearly, the effect may be general, specific, or both, depending on the way the experiment is arranged and the measurements are made. In a conditioning experiment, for example, in which punishment is employed for the purpose of establishing a discrimination, the effect is of necessity highly specific. A response to, say, one tonal frequency is reinforced, and the responses to other frequencies are weakened or eliminated. It would be obviously nonsensical to expect unpunished responses to be affected in the same way as punished ones. At the other extreme of the continuum of specificity-generality is a situation in which punishment is administered without reference to any particular response, say, after a block of trials. In that case, the only possible effect that punishment *can* have is general, inducing an attitude of caution, serving as an incentive, etc. The general conclusion on the generality-specificity issue is that punishment has both general and specific effects, sometimes one, sometimes the other, sometimes both, depending on its role in the total experimental situation.

What is the mechanism by which punishment exerts its influence? There has been considerable speculation about the events in the organism which mediate the effect of punishment. This is, of course, an example of the search for intervening variables. Punishment is the independent variable, the effects of punishment (positive and negative) are the dependent variables. In terms of what hypothetical events can the relationship between these two variables be conceptualized? The answers which investigators have suggested fall into three broad classes: (1) mechanisms of action which are held to be specific to punishment, (2) explanations which reduce whatever effect punishment may have to the ultimate action of rewards, and (3) explanations which ascribe the effectiveness of punishment to the information or perceptual emphasis which it provides.

1. *Mechanisms specific to punishment.* Throughout his writings on punishment Thorndike has referred to the weakening or stamping out of stimulus-response bonds. Such weakening effects as he has found were ascribed to processes detrimental to the conductivity of the neurones. For the last fifteen years, however, he has argued against significant effects of punishment and therefore has not further concerned himself with a mechanism to account for the weakening of specific S-R bonds by punishment, putting the entire burden of explaining learning on the action of rewards.

An important suggestion as to the mechanism by which punishment affects the emission of specific responses comes from students of operant conditioning (68, 237). Their experiments deal with the effects of negative reinforcement (punishment) on the lever pressing behavior of rats. The general conclusion is that punishment results in a *suppression* rather than *weakening* of a response. Administration of punishment temporarily depresses the rate at which responses are emitted, but the suppressed behavior will be released when circumstances

are more favorable. In terms of Skinner's concept of reserve, punishment affects the relation between reserve and rate of emission of responses but has no effect upon the reserve itself. The effect of punishment is therefore emotional, leading to a temporary depression of function but not to a permanent weakening or elimination of responses. These findings not only serve to emphasize the differential effects of reward and punishment on the emission of responses but also convey an important methodological warning: the strength of a response at the end of a period of punishment cannot be used as a reliable index of the true long-range strength of the response (68). Punishment is not a mechanical stamping-out process. Its effect must be considered in relation to the experimental situation.

The *emotional* effect of punishment is also emphasized in Guthrie's attempt to account for the elimination of wrong responses (90). For Guthrie, the fundamental condition of all learning is conditioning by contiguity in time. It is *movements*, not acts, which are conditioned to stimulus cues. The more aroused the organism, i.e. the greater the emotional excitement, the more varied and intense the movements which occur. These varied and intense movements produce a correspondingly high degree of proprioceptive stimulation. Proprioceptive stimuli are important cues to which responses are conditioned. Hence punishment facilitates conditioning (of withdrawal responses, for example) not because punishment is annoying but because punishment arouses the organism and provides a multiplicity of cues for the establishment of conditioned responses. Guthrie insists that the action of punishment cannot be in any way clarified by referring to the action of annoyers. Suppose we define annoyance as a state of affairs which the animal avoids or changes. "But this ability to avoid is just what is necessary to explain unless we assume that learning has already taken place. If annoyance means avoidance, we had to learn to be annoyed at annoyers (by emotional reinforcement). It is what the punishment makes an organism do that counts not what it makes him feel!" (90, p. 14).

From what may be called stimulus and/or response explanations of punishment we turn now to explanations which put the main emphasis on the changes wrought by punishment in the subject's approach to the learning situation and his general manner of performance. We are dealing here with qualitative descriptions of performance under punishment. Several investigators have stressed the attitude of *caution* which the subject assumes when he is punished for incorrect responses. The learner is circumspect, he steps warily (10, 24, 25, 26, 172, 173, 218). Muenzinger has found that a punishment or obstacle makes the subject *pause* before he proceeds with his response (196, 200, 202, 203). During this pause he is exposed to relevant stimulus cues. In this regard shock and an enforced delay (200, 203) or an obstacle such as a gap (202) after the choice-point are equivalent. According to Honzik and Tolman (116) a state of heightened vigilance may make this increased exposure to stimulus cues especially effective. Thus punishment is described as making the organism more sensitive and vigilant, and as enhancing the animal's opportunity to learn about the situation and indulge in successful vicarious trial-and-error. Similar qualitative observations describe punishment as raising the general incentive level (57, 79, 80).

2. *Explanation of effective punishment in terms of reward.* When Thorndike jettisoned punishment as a condition of learning, he left open the possibility that it may have an indirect effectiveness by speeding up the occurrence of acts

which are rewarded. Thus *punishment may be a condition for the eventual application of reward* and in this way a determinant of learning, though one step removed from the locus of the critical event. Hull's theory of primary reinforcement which makes all learning contingent on need-reduction clearly implies such an approach to punishment. The reduction of the effect of punishment to the action of reward has been made explicit in the writings of Mowrer (181, 183, 184, 185, 186, 188). According to the conception which Mowrer presents, *all motives involve tension or discomfort*. Thus there is discomfort or tension due to hunger just as much as there is discomfort or tension due to shock. In this sense, all learning, dependent as it is on motivation, involves punishment, the punishment inherent in the motivating tension. Given the motivation (tension), learning occurs when the tension is reduced. Tension may be reduced either through the occurrence of a consummatory response (e.g., reduction of hunger by feeding) or through *escape from punishment* (e.g., escape from shock). In Mowrer's own words, "... it is therefore meaningless to say that one type of learning is 'through reward' and another type is 'through punishment'; each is an essential aspect of a single dynamic process" (184, p. 424). Similarly Muenzinger (199) argues that it is "poor logic and bad science" to compare learning by reward and learning by punishment. Without a primary state of imbalance and its subsequent reduction there can be no learning. Punishment is merely a way of increasing this imbalance or tension in order to make the subsequent need reduction more effective as a reinforcing state of affairs.

Such a conceptualization assigns to punishment a much more important systematic role in learning than does Thorndike's. For Thorndike, the effects of punishment are variable and unpredictable. Punishment has no vital place in his picture of the learning process. In the view represented by Mowrer and Muenzinger, on the other hand, punishment is at the very fountainhead of learning. For punishment induces tension, and without tension the behavior which leads to learning by reinforcement would not occur. Thus the law of effect again becomes a law of reward and punishment. It has, however, undergone a crucial transformation. Formerly *either reward or punishment* was believed to be a determinant of learning, and punishment was, as it were, reward with a negative sign attached to it. The modern version is a law of reward *after* punishment, for punishment provides the stimulus which makes a reward a reward. Both reward and punishment are integral parts of the learning situation, provided they occur in the proper sequence and are functionally connected with each other. This analysis of the role of punishment, which Mowrer has defended on the basis of his experimental work, has all the beauty and all the dangers of simplicity. Here is an integrated, coherent account of learning, *granted the basic assumption that learning cannot occur without initial tension and subsequent tension-reduction*. It is precisely the granting of this assumption which is one of the controversies with which modern learning theory abounds. Thus we have come in our discussion to the same parting of the ways to which our discussion of reward has led us. In which direction progress lies depends on whether or not tension and tension-reduction can stand up as necessary (though not sufficient) conditions of learning.

3. *Punishment or emphasis?* In general, an individual is subjected to punishment only when he has committed a wrong. In learning experiments, the same procedure has usually been followed. The subject is punished when he commits an error. Sometimes it follows unequivocally from the nature of the situation what

constitutes an error (e.g., blind alley in a maze), at other times the experimenter decides arbitrarily which responses are to be considered right and which ones wrong (e.g., in learning word-number combinations in a Thorndikian experiment). Thus errors and punishment have been associated in investigations of the law of effect until Tolman, Hall and Bretnall (294) decided to explore the effect of punishment for *right* as well as for wrong responses. If punishment does, indeed, weaken associative connections, then punishment for correct responses should slow up the learning process, whereas punishment for wrong responses should result in faster learning. These expectations were not confirmed. Tolman, Hall, and Bretnall found that (1) adding a shock to an auditory signal for right responses did not slow up learning significantly and (2) adding a shock to an auditory signal for wrong responses did not speed up learning but on the contrary slowed it down. Reinforcement of right responses was always more effective than reinforcement of wrong responses. These findings were so strikingly at variance with what the traditional law of effect would have predicted that the article reporting these findings was entitled "A disproof of the law of effect."¹²

Tolman, Hall, and Bretnall believed that the reinforcing stimuli—both auditory signal and shock—do not stamp in or stamp out responses but rather serve to *emphasize* (perceptually or cognitively) the reinforced responses. Moreover, an emphasis upon correct responses favors learning whereas emphasis on wrong responses does not. Any relatively violent emotional stimulus may in addition have disruptive effects which will counteract the favorable influence of emphasis. Other experimenters took up the idea of administering punishment for right as well as wrong responses. The work of Muenzinger has already been referred to. In the first of his series of experiments he found that punishment for right responses and punishment for wrong responses were equally effective (193). In a subsequent repetition of his experiment he found shock for wrong responses inferior (196). In one experimental variation, shock was administered for correct choice, incorrect choice, and at the time of feeding. All three conditions accelerate learning significantly (57). As in other studies of punishment the intensity of the shock relative to the task is a crucial parameter (87).

The experiment of Tolman, Hall, and Bretnall was of considerable systematic significance because it shifted attention from punishment as an automatic mechanism to punishment as a perceptual event. The learning task may be regarded as a problem which the learner tries to solve. Anything which helps the learner to distinguish correct from incorrect responses facilitates learning. A punishment may merely serve to designate a given response as either right or wrong at the pleasure of the experimenter.

INFORMATION VERSUS EFFECT

The efficacy of punishment for right as well as wrong responses is closely linked to a general problem which has been repeatedly raised in

¹² Goodenough (84) criticized Tolman's findings on statistical grounds but later experiments have tended to confirm Tolman's results (119, 214, 234).

discussions of the law of effect. Can the efficacy of rewards and punishments be explained in terms of the *information* which they convey to the subject about the correctness or incorrectness of responses? If the effect of reinforcement were always mediated by information, it would not be possible to speak of a direct effect of satisfaction (need-reduction) or annoyance on stimulus-response connections. Rather, the hypothetical sequence of events in the organism would then become: stimulus-response-reward or punishment-information-strengthening (or weakening) of the association. In such a description, the mechanism by which information leads to the strengthening of associative connections would be left open. One could, like Tolman, think of the formation of sign-gestalt expectations (288, 291) or assert that information leads to tension-reduction and that learning through information is only a special case of the law of effect. The problem of effect versus information has not always been posed in all-or-none terms. Granting that both factors may be operative, investigators have been interested in gauging their *relative* importance.

Information is a concept which does not easily lend itself to experimental analysis. There is always the vexing possibility that what the experimenter considers to be mere information may act as reward, and conversely that what the experimenter believes to be a reward (need reduction) may serve as a source of information. There may, moreover, always be self-administered information (229). Nevertheless, experimental situations have been designed which at least made it possible to speculate about the role of information in the learning process. These experiments have been concerned with (1) clearly non-informative after-effects, (2) the question whether a learner need necessarily be aware of what he is learning, (3) the relevance of after-effects to the connections which they strengthen.

Non-informative after-effects. In our discussion of punishment experiments were cited which show that punishment administered without reference to a specific response facilitates learning. Similarly, rewards given after a block of trials and not following a particular correct response have been found effective (60, 281). In general, however, effects which give specific information have been found superior to non-informative ones (60, 224, 298).

The study of "neutral" after-effects has a direct bearing on the problem of effect versus information. An after-effect is "neutral" if it neither rewards, punishes, nor informs. Nonsense words, meaningful words, flashes of light, and clicks have been used as after-effects which merely happen after an S-R connection, presumably without telling the subject anything about the correctness of his response. Several investigators

have found that such after-effects strengthen responses to some extent (165, 166, 255). The very ambiguity of such after-effects has made it possible to interpret them in different ways. Thorndike and his co-workers believe that ambiguous after-effects *may* be interpreted as rewarding by the subject and hence function as rewards (165, 166). Stephens (255), on the other hand, suggested that the mere fact that *something* happens after a response, that the connection is "attended to" facilitates learning. Something happening after a response may, however, be interpreted by the subject as conveying information. "Neutral" after-effects have repeatedly failed to influence learning when the factor of information was carefully controlled (37, 38, 298).

It seems that the use of neutral after-effects is a highly unsatisfactory procedure. The fact that an experimenter considers an after-effect neutral does not, of course, guarantee that the subject interprets the event in the same way. It seems puzzling that some experimenters first call an effect neutral, then turn the tables on themselves and assert that it was not neutral after all but rewarding. If it is neutral after-effects that we want to study, we should make sure that the effects are indeed neutral and then consider them neutral throughout. On the other hand, if we are interested in the effect of rewards, the study of ambiguously neutral after-effects provides a very devious avenue of approach to the problem.

Awareness of what is being learned. If learning by reward can take place while the learner is not aware of what he is learning, then certainly effect and information could not be considered coextensive. Thorndike repeatedly claimed that rewards are effective even if there is little or no opportunity for inner rehearsal of the right responses (266, 269, 275), though he has been contradicted on that point (37, 298). Thorndike and Rock (282) then offered what they consider conclusive proof that learning without awareness of what is being learned can, indeed, take place.¹³ In their experiment, learning depended on the *discovery of a principle* by the subjects. Yet there was only gradual improvement under reward and no sudden increase in successful responses. It seemed that subjects achieved insight without knowing that they did. The interpretation of this experiment depends entirely on the assumption that gradual improvement indicates lack of awareness. It turned out, however, that even subjects who are *explicitly taught* the principle on which their learning depends still may show gradual improvement (131). The point here is that understanding a principle and using it are two different things. Slow improvement may reflect lack of insight but also may be due to the learner's inability to translate his understanding into smooth, efficient action. This attempt to prove learning without awareness must be considered doubtful. On the other hand, positive evidence

¹³ For other experiments demonstrating that subjects can learn without being aware of what they learn, see Thorndike (269, 276).

has been offered for the importance of awareness (104) and knowledge of results (7, 82, 140, 225) for successful learning. Whether or not learning without awareness of what is being learned can take place must remain an open issue.

The role of relevance. A satisfying after-effect, according to Thorndike, does not act logically but in a mechanical way. It is likened to a natural force applied to do work and need not, therefore, be *relevant* to the activity of the organism at the time at which the reward is given. Suppose a subject is engaged in activity directed toward goal A. In the course of his trial-and-error behavior he is given a reward which leads not to goal A but to an irrelevant goal B. Nevertheless the response which led to the reward is reinforced. Such findings have been reported for both human and animal subjects (164, 271, 274, 276), casting doubt on the theory that rewards function primarily as sources of information relevant to the learning task. How conclusive are such experiments on the role of relevance? Relevance and irrelevance are terms which refer to the *experimenter's* interpretation of the situation. The subject's interpretation and the wants under which he operates are an altogether different matter. The only conclusion which can be drawn from such experiments is that responses which occur in temporal proximity to a reinforcing state of affairs are strengthened (Hull). Relevance is a normative term reflecting the experimenter's judgment.

Even if we agree that relevance is a useful dimension for the description of reward, the evidence on the role of relevance remains inconclusive. There are strong indications that, at least in some experimental situations, rewards irrelevant to the subject's motivation fail to facilitate learning. The work of Wallach and Henle (305, 306) is a case in point. In a typical Thorndike situation (learning word-number combinations) the subjects were informed that they were participating in an experiment on extra-sensory perception and that responses called right on a given trial might or might not prove correct on subsequent occasions. These instructions rendered the rewards irrelevant to the subject's task—guessing numbers by ESP. As a result, rewarded responses were not repeated more frequently than punished ones, nor did these experimenters obtain a level of repetition of wrong responses at all comparable to that reported by Thorndike. Wallach and Henle conclude that "with the learning motive eliminated no automatic effect of reward seems demonstrable." Their argument is further strengthened by the fact that a change in instructions, telling the subject that responses would no longer vary in a random fashion caused a highly significant increase in the number of rewarded responses that were repeated.

The results reported by Wallach and Henle strongly suggest that rewards may not act as illogically and blindly as Thorndike's findings had indicated. The subject's *attitude* toward the learning situation needs to be taken into account. In the absence of explicit instructions not to

learn, subjects will by and large instruct themselves to learn and utilize whatever cues the situation provides. A reward, even an irrelevant reward, is such a cue because a rewarded response is different from the majority of responses which remain unrewarded. Thus there is a tendency to follow rewards and to consider them relevant if not dissuaded by instructions to the contrary. Wallach and Henle have clearly shown that a change in attitude can resist the "mechanical" impact of a reward. Investigators of human conditioning have long recognized the importance of subtle attitudinal factors (108). Attitudinal factors are probably of equal importance in the study of S-R connections by Thorndike's technique.

The issue of information *versus* effect remains unsettled. The difficulty is that it has not been possible to prove conclusively that rewards do not yield information even when experimenters hope and believe that they do not. An experimenter provides what he considers a non-informative reward but the subject may extract information from it, especially if he is a subject capable of symbolization.

THE SPREAD OF EFFECT

There is one phenomenon which more than any other has bolstered the view that rewards act mechanically and blindly. This phenomenon is the spread of effect. Again the pioneer investigation was carried out by Thorndike. In 1933 he published evidence that a reward strengthens not only the connection which it directly follows and to which it belongs but also the connections which *precede and follow* the rewarded response (270). The closer in the series an item is to the rewarded connection the more it benefits from this spread of effect. Thus there is a double (*before and after*) gradient of effect, with the items preceding the reward showing somewhat less strengthening than those following the reward. Degree of spread depends primarily on the number of serial steps separating an item from the reward and not so much on sheer temporal proximity.

Thorndike considered this finding an independent proof of the law of effect (272). The discovery of the spread phenomenon gave Thorndike new confidence not only on backward action but also in the mechanical action of satisfiers. Again in his own words:

The satisfier acts (upon a neighboring punished connection) unconsciously and directly, much as sunlight acts upon plants, or an electric current upon a neighboring current, or the earth upon the moon. From a satisfier issues a strengthening force which the connections absorb. Which of them will absorb it depends on the laws of nature not of logic and teleology (270, p. 48).

As to the particular mechanism responsible for the spread, Thorndike entertained two hypotheses: (1) *the scatter hypothesis*: the strengthening effect, being not logical but a biological force will sometimes miss its mark, striking preceding and succeeding connections. The gradient thus represents the decreasing probability of chance errors in the action of the satisfier. (2) *The spread hypothesis*: the confirmatory reaction caused by the reward may be diffuse, spreading out its influence over a range of items.

Thorndike's findings have been repeatedly confirmed with both human and animal subjects (15, 72, 134, 135, 136, 197) and the phenomenon became known as the "Thorndike effect." Agreement ceases, however, when it comes to an interpretation of the phenomenon. Discussion has centered around a number of related questions: (1) Is the gradient of effect really double-winged or is the *before* gradient spurious? (2) Is the gradient of effect a gradient of variability? (3) Is the gradient a result of perceptual emphasis on the rewarded response rather than of the mechanical action of satisfiers? (4) Is the gradient of effect really a gradient of response habits?

1. *Is the gradient of effect bidirectional?* For Thorndike one of the most important characteristics of the spread of effect is its bidirectionality: rewards exert their influence both in the forward and backward direction. The interpretation of the *before* (backward) gradient runs into a methodological difficulty, however. A connection which precedes a reward *also* follows a reward. In Thorndike's experiments rewards often followed each other rather closely. In criticizing Thorndike's data Tilton therefore raised the question whether the backward gradient might not be a spurious function of a large forward one (286). He also pointed out that serial position needs to be taken into account with this type of material. Putting these considerations to the test—taking serial position into account and attempting to eliminate the influence of the forward gradient from the backward gradient—Tilton still found a double gradient but from *both success and failure*. In the case of the *failure* gradients, the backward one is more pronounced, while the forward gradient is the more pronounced in the case of success. Other investigators also report the existence of a failure gradient (197, 332). The double-winged gradient is thus confirmed, it is true, but we find that the effect of *Wrong* spreads as well as the effect of *Right*. These results render doubtful Thorndike's theory of the scatter or spread of the confirming reaction. As an alternative explanation Tilton suggests that connections which the experimenter regards as discrete are not functionally separate. Items in a series may be sufficiently unified to be affected by success and failure as a unit (286). It is not entirely clear, however, why such "sequential unity" should manifest itself as a double

gradient. Further doubt is cast on this hypothesis by Zirkle's finding that the rewarded item and the adjacent punished ones need not be qualitatively similar in order for the Thorndike effect to appear (331). Even more strikingly, Zirkle has demonstrated that it is the *response* adjacent to the rewarded item which is strengthened, not the adjacent S-R connection. When the relative order of wrong series items about a right item is shifted from one presentation to the next, there is no Thorndike effect when repetitions of S-R connections are counted. However, a clear-cut Thorndike effect appears when repetitions of *responses* are counted by *step-position* alone without regard to the shifting positions of the stimulus items. It has also been found that reward strengthens the *early* wrong responses to near-by stimuli not merely the last response (174).

2. *Spread of effect as a gradient of variability*. There have been other reformulations of the Thorndike effect. Muenzinger and Dove (197) disposed of the scatter hypothesis which ascribes the spread of effect to inaccurate placement of reward and consequent uncertainty of recall. A gradient is present, with a steeper slope than usual, even if the successful response is learned beforehand so that there can be no possible confusion between the correct response and the surrounding wrong ones. For a gradient of uncertainty Muenzinger and Dove substitute a gradient of uniformity or variability. Success produces a gradient of uniformity: this is in essence the original Thorndike effect. Failure produces a gradient of variability: not only does the wrong response itself tend to be varied but right responses near the wrong one are not repeated with the same degree of uniformity as those farther away in the series. Spread of variability under punishment is also reported by Stone (260). This analysis still leaves open the question as to what intervening mechanism the changes in variability should be ascribed.

3. *Is the effect gradient due to perceptual organization?* The systematic differences in approach to the problem of effect in general are necessarily reflected in interpretations of the spread of effect. Again we find those who regard rewards and punishments primarily as perceptual events opposing those who think of the mechanical action of satisfiers and annoyers. The conditions under which the Thorndike effect generally appears are especially favorable to an analysis in perceptual terms. The rewarded item is an isolated *Right* in a long homogeneous series of *Wrongs*. The announcement of *Right* may be regarded as a figure against a homogeneous ground of *Wrongs*. Wallach and Henle were the first to suggest that it is the extreme *crowding* of items which interferes with memory for wrong responses and contributes to the subject's tendency to repeat them (305, 306). The relation between perceptual isolation and the Thorndike effect was systematically investigated by Zirkle (332). He found a clear-cut positive relation between the degree of isolation of a rewarded response and the steepness of the effect gradient

(especially the *after gradient*). He also found a failure gradient around a perceptually isolated wrong response. On the basis of his experimental results Zirkle proposes a *theory of isolation* to account for the Thorndike effect. "A satisfier isolates tendencies which are at hand when it happens . . . responses neighboring upon a key response *tend to become isolated themselves* because of their association with the key response" (332, p. 312f.). A theory of isolation is, of course, a perceptual theory of the spread of effect. That which stands out in a homogeneous series becomes a focus of retention (similar to the von Restorff phenomenon). This interpretation is closely akin to Tolman's law of emphasis. Thorndike and Zirkle thus clearly represent the opposition between an interpretation of the effect gradient in terms of "hit-and-miss" mechanical reinforcement on the one hand and a view stressing the laws of perceptual organization on the other.

4. *Is spread of effect a gradient of response habits?* In terms of an orthodox connectionist analysis the Thorndike effect indicates that the influence of reward spreads to neighboring *S-R connections*. It is the S-R connection that is the basic unit of analysis. It is possible, however, to account for at least part of the effect gradient by an analysis of the *response* sequence, disregarding as it were the stimulus side of the S-R connection. We have already referred to Zirkle's finding that a Thorndike effect can be demonstrated when responses are counted by step-position even though the stimulus order is changed from presentation to presentation. The problem of response habits was systematically attacked by Jenkins and Sheffield (136). They found that (1) when the rewarded response itself was not repeated few adjacent errors were repeated and the Thorndike effect failed to appear; (2) when the rewarded response itself was repeated it was accompanied by a high level of repetition of errors and a typical effect gradient. Thus repetition of the rewarded response appears to be a necessary condition for the appearance of the Thorndike effect. The spread of effect may therefore be due not to the automatic action of reward but to the subject's tendency to repeat the same sequence of responses from trial to trial, i.e. the subject's "guessing habits." Responses are not independent of each other: choice of a response depends on the preceding responses. A repeated rewarded response ensures that "the errors following reward will be frequently preceded by the same response, i.e., will have a common antecedent stimulus." It is interesting to recall in this connection that preliminary rehearsal of the rewarded response yields an especially steep effect gradient (197) and that, on the other hand, a gradient fails to appear when the subject is not motivated to repeat the rewarded response (305, 306). Guessing sequences have also been reported with other types of multiple-choice responses (72), and there seems little doubt that at least part of the Thorndike effect can be analyzed in these terms. The demonstration of guessing habits does not necessarily dis-

prove the automatic spread of reward as conceived by Thorndike. It may well be that both some "stamping-in" mechanism and response habits interact in the production of the Thorndike effect.

As in the case of other conditions of reinforcement, it is necessary to stress the dependence of the spread of effect on the parameters of the experimental situation. As we have seen, the Thorndike effect fails to appear when subjects lack the motivation to learn (305, 306) though distraction does not seem to affect it (136). In animal subjects, increase in drive results in a higher frequency of repeated responses in general, with a consequent flattening of the after-gradient (134). Similar results are obtained with an increase in incentive (135). Under such conditions, only responses *preceding* the reward yield a statistically reliable gradient. Such delicate dependence of the gradient on particular experimental conditions counsels against hasty analogies between the action of satisfiers and the "action of the earth upon the moon."

PARAMETRIC STUDIES

We have repeatedly stressed the functional dependence of the effects of reinforcement on the parameters of the experimental situation. We now turn to a consideration of studies that are primarily concerned with the influence of such parameters of reinforcement. No exhaustive survey will be attempted but the major types of functional relationships will be illustrated.

Amount of Reinforcement

Thorndike's original statement of the law of effect included the assertion that "the greater the satisfaction or discomfort the greater the strengthening or weakening of the bond." Demonstration of the correlation between amount of reinforcement and degree of learning has proved much easier with animal subjects than with humans. The motivation of animal subjects is more easily controlled and hence quantitative variations in reward are more effective. Increasing the food ration of a hungry rat affects behavior more drastically than adding 0.4 cent to the announcement of *Right* in a Thorndike experiment.

Turning to animal studies first, there are a number of experiments with different species and different learning tasks which support the generalization that increases in reward lead to improvements in learning and performance. Chickens (86) as well as rats (42) run down a runway at a faster speed when the amount of food reward is increased. Increases in reward also cause chicks to learn a maze with fewer errors (320), rats to show greater resistance to the extinction of a conditioned

response (77), and chimpanzees to tolerate longer delays between the presentation of the stimulus and the response (208). Although the functional relationship between amount of incentive and performance varies from situation to situation, it is clear that the relationship is not linear. Equal increments in incentive do not lead to equal increments in performance. Thus Crespi, who was especially concerned with the quantitative relation between amount of incentive and performance reports a sigmoid relationship over a wide range of variation (42). Hull, on the other hand, believes that the relationship is best represented by a simple positive growth function (127).

The mechanism by which increments in reward exert their influence is still under discussion. Hull has suggested that an increase in the amount of reinforcement raises the limit to which the curve of habit strength approaches as an asymptote, although the rate of approach may be constant for all amounts of reinforcement. Hence increases in the amount of the reinforcing agent result in greater increments of habit strength per reinforcement. Hull also points out that in a conditioning situation the reward provides an important component of the stimulus situation which is conditioned to the response being reinforced. In such a situation a large reward stimulus evokes stronger, more vigorous, and more persistent responses than a small reward stimulus. Thus an increment in reward not only increases the amount of consummatory activity (need reduction) but also provides a more distinctive cue to the animal being conditioned (127). In this connection it is important to distinguish between sheer physical amount of reward and the amount of activity involved in the consumption of the reward. Wolfe and Kaplon were able to show that amount of reward and amount of consummatory activity are experimentally separable conditions of learning. Learning improves as a function of sheer amount of reward: a whole kernel of corn is a more effective incentive for chickens learning a maze than one-fourth of a kernel. However, learning also improves if the amount of consummatory activity is increased for a constant reward: when the whole kernel of corn is divided into four separate quarters the maze is learned better than with the equivalent amount given in one piece. Of the two factors, amount of consummatory activity influences learning to a greater extent than sheer amount of reward (320).

The role played by amount of reinforcement has also been explained in terms of the subject's attitude rather than in terms of differential strengthening of S-R connections. Crespi has proposed a two-factor theory of *incentive-value*. He conceives of incentive-value as proportional to the distance between the subject's level of expectation (both of quantity and quality of reward) and the level of attainment. Attainment which does not reach the level of expectation is frustrating and affects learning adversely, while attainment above expectation is "elating"

and improves learning. It is not the sheer amount of reward alone which is decisive nor the animal's expectancy alone but rather the relation between the two. A striking illustration of this principle is the fact that rats will perform significantly better with no incentive at all than with a very small incentive. A very small incentive serves to whet the animal's appetite, raises his level of expectation and eventually leads to frustration. In accordance with the same principle, downward shifts in the amount of incentive lead to poorer and more variable performance whereas upward shifts improve the performance (42, 42a). Crespi's conceptualization is reminiscent of the experiments on *level of aspiration* of human subjects. That the effectiveness of a given quantity of reward cannot be evaluated without reference to the subject's expectancy is also stressed by Cowles and Nissen (40) on the basis of experiments on delayed responses in chimpanzees. Similarly Nissen and Elder (208) report that increases and decreases in amount of incentive affect not only the response within a given trial but succeeding trials as well. Such perseverative effects suggest the operation of reward expectancy.

When human subjects are used, comparisons of different amounts of reinforcement yield variable and inconclusive results. Slight increases in rate of learning as a function of increases in reward—addition of small money gains to the announcement of *Right*—have been found in some Thorndikian experiments (60, 223, 280). The effects are exceedingly slight if the amount of reward varies within the same series (223) and are somewhat more pronounced if incentives are changed from series to series (280). On the other hand, promise of a reward was found as effective as actual administration of the reward (78). In situations of this type, the experimenter never knows whether an increase in reward is experienced as such by the subject. Moreover, if different rewards are presented in the course of an experiment, their effects may interact with each other. Whatever effects there are, may possibly not stamp in individual responses more firmly but rather affect the general level of the subject's motivation and thus affect learning secondarily (171).

Frequency and Pattern of Reinforcement

The simplest hypothesis relating frequency of reinforcement with strength of learning is that each reinforcement adds to the strength of the S-R connection being reinforced. This hypothesis is plausible, at least at first blush, in the light of empirical data. The percentage of conditioned responses usually increases as a function of the number of reinforcements (108), human rote learning steadily improves on successive trials (171). Indeed, it is difficult to think of learning situations in which an increase in number of trials does not result in better perform-

ance.¹⁴ Frequency *per se* cannot be profitably considered a significant condition of learning. Rather, repeated pairings of stimulus and response allow effective conditions of learning, (such as simultaneous conditioning, reinforcement by need reduction, confirmation of expectations, etc.) to exert their effects. There would probably be little disagreement with such a general formulation of the role of frequency. Agreement ceases when the assertion is made that each individual reinforcement in a series of reinforcements contributes a differential increment (say, ΔgH_R in Hull's language) to the existing habit strength. The theoretical question at issue is whether or not the effects of successive reinforcements are additive and cumulative. Is the law of effect a law of cumulative effect?

The assumption that the effects of successive reinforcements continuously cumulate in time is central to Hull's theoretical system (127). He conceives of habit strength as a monotonic increasing function of the number of reinforcements. As habit strength approaches the subject's physiological limit, the increment from each reinforcement progressively decreases in magnitude. Thus habit strength is a simple positive growth function of the number of reinforcements. Habit strength, as Hull uses the term, is a theoretical construct which can be measured only indirectly through its behavioral manifestations. There are a number of behavioral studies whose results are consistent with Hull's quantitative picture of the growth of habit strength:

1. Reaction amplitude increases as a function of the number of reinforcements (118).
2. Reaction latency decreases with the number of reinforcements (235).
3. The number of trials required for experimental extinction may be proportional to the number of reinforcements (209, 317, 329).
4. The more frequently a response has been reinforced the greater the probability that the appropriate stimulation will evoke that response (162, 269, 276, 329).

Unfortunately not all experimental results available fit into Hull's theoretical picture. Leeper (160) has criticized Hull for what he considers a biased selection of illustrative experiments. There is substantial experimental evidence that partial reinforcement, i.e., reinforcement on only a fraction of the trials rather than on each trial may be at least as effective as continuous reinforcement. Thus Humphreys showed that reinforcement on 50 percent of the trials is as effective in the establishment of a conditioned response as reinforcement on 100 percent of the

¹⁴ A notable exception is Skinner's finding that a single reinforcement results in a series of lever pressings by the rat at an optimal rate (237). One reinforcement is sufficient to establish an adequate response strength. Further reinforcements serve to build up a reflex reserve, which is subsequently emptied in the absence of reinforcement (extinction).

trials and that partial reinforcement may result in *greater* resistance to extinction than continuous reinforcement (128, 129). Humphreys believes that conditioned responses occur to the extent that the subject *expects* the reinforcing stimulus to follow the conditioned stimulus, extinction occurs to the extent that the subject no longer expects reinforcement. Partial reinforcement during the initial training period makes it difficult for the subject to shift from expectation of reinforcement to expectation of non-reinforcement. In terms of the expectancy hypothesis it is necessary clearly to distinguish between frequency of trials and frequency of reinforcements as conditions of learning. The relation between these two frequencies determines the subject's expectations and hence the course of acquisition and extinction of responses (130). The equal effectiveness of partial and continuous reinforcement has, however, also been ascribed to the influence of secondary reinforcement during the ostensibly unreinforced trials (52). Such an explanation would be in conformance with a description of the reinforcement process as continuous and cumulative. The relationship between number of trials and relative frequency of reinforcement is complicated by the phenomenon which Hovland has called "inhibition of reinforcement" (117). A massing of reinforcements results in a weakening of conditioned responses, with spontaneous recovery within a short interval of time. Inhibition of reinforcement as well as expectancy have to be taken into account in the analysis of partial reinforcement (75, 76).

The studies cited thus far by no means exhaust the evidence for the equal effectiveness of partial and continuous reinforcement. Brogden (16) showed that conditioned flexion to shock was established as readily with 40 percent reinforcement as with 100 percent. Decrease in the frequency of reinforcement had the positive effect of eliminating a great deal of the animal's diffuse and restless behavior. When food rather than shock was used as the reinforcing stimulus substantially the same results were obtained, except that there was a slight decrease in the frequency of response (CRs on $\frac{4}{5}$ of the trials) when reinforcement was applied only 20 percent of the time. Brogden concludes that application of the reinforcing stimulus serves primarily as an *incentive* to the subject, and thus a low frequency of reinforcement may maintain the conditioned response at a high level.

Mowrer and Jones (187), who confirmed Humphreys' findings, present an argument reconciling these results with the law of effect. The effects of a reward need not necessarily be restricted to the particular response that occurs just before the reinforcement. The reinforcement applies to preceding responses as well (though to a decreasing extent). If we think in terms of *response units* (sequences) each of which is followed by a reward rather than in terms of individual responses which sometimes are reinforced and sometimes are not, the apparent advantage of intermittent reinforcement disappears. On the contrary, in

terms of a response-unit analysis, the intermittently reinforced group which has to expend more effort in order to obtain a reward gives fewer extinction responses.¹⁵ The response-unit hypothesis of Mowrer and Jones points to the importance of considering the *temporal pattern* of reinforcement and not only the sheer frequency. Skinner's results with periodic reconditioning are a relevant case in point (237). When the lever-pressing response of the rat is reinforced periodically, e.g., every three minutes, the animal's rate of response not only tends to become uniform, but also the more frequent the reinforcement the more rapid is the rate. On the other hand, when reinforcement is at a fixed ratio, i.e., when the final member of a fixed number of responses is reinforced, the *less* frequent the reinforcements the higher is the rate of response. Such laws of operant behavior can be analyzed only in terms of the total temporal pattern of responses and reinforcements (e.g., Skinner's concept of reflex reserve). The importance of the temporal pattern is also illustrated by Brunswik's finding that in discrimination learning *probability of success* is an important determinant of the animal's behavior at a choice point (21).

In summary, it is clear that frequency of reinforcement is an important determinant of the strength of learning. However, partial reinforcement can be as effective as, and more effective than, continuous reinforcement. Reinforcements are not always simple additive units, and the temporal *pattern* of a series of responses and reinforcements gives rise to behaviors which cannot be predicted in terms of a simple monotonic relationship between frequency of reinforcement and strength of learning.

Delay of Reward and Gradient of Reinforcement

The degree to which reinforcement strengthens a response depends in part on the time interval that elapses between response and reinforcement. Experimental inquiry has been directed at two interdependent problems: (1) the difference between immediate reinforcement and delayed reinforcement and (2) the exact quantitative relationship between length of delay and strength of learning, i.e., the nature of the *gradient of reinforcement*.

Thorndike early suggested that satisfying states of affairs are the more effective the closer they are temporally to the S-R bond. "Other things being equal, the same degrees of satisfyingness will act more strongly on a bond made two seconds previously than on one made two minutes previously . . ." (265, p. 172). A few of the early investigations of delayed reward failed to bear out this prediction (309, 313), but it is now clear that in these experiments secondary reward was not con-

¹⁵ Mowrer and Jones (187) also point out that the law of effect is compatible with Humphreys' expectancy hypothesis. The fulfillment of an expectation may serve as a reward (need reduction).

trolled—the animal subjects were delayed in the *food chamber* of a maze—so that there was no effective delay of reinforcement. When these first attempts are discounted, the experimental literature shows general agreement on the superiority of immediate over delayed reward. The detrimental effects of delayed reinforcement have been demonstrated with animal subjects in maze learning (34, 100, 318), problem box learning (222), the formation of discrimination habits (35, 321), and in the establishment and maintenance of operant responses (210, 211, 237).¹⁶ Similar findings have been reported for delayed punishment (32, 141, 192, 308, 311, 321, 323).

Different investigators used different periods of delay but there were strong indications that the most serious detrimental effects were concentrated in the first minute of delay. Formal quantification of the functional relationship between length of delay and strength of learning are found in the writings of Hull (127) and Perin (210, 211). On the basis of theoretical calculations, which fit a considerable amount of empirical data, Hull concludes that (1) habit strength is a negative growth function of the time separating the response from the reinforcement and (2) the asymptote of this gradient is zero, i.e., with a sufficiently long delay reinforcement becomes ineffective. In the case of the rat, the gradient reaches zero at a delay of about 30 seconds (210). This rather short gradient of habit strength constitutes the *gradient of reinforcement*.¹⁷ It is important to bear in mind that the gradient of reinforcement refers to the effects of different intervals *between single responses* and that it does not apply to one of a series of responses leading to a common reward (108). In the case of a series of responses, e.g., in a maze, the gradient is more extended and more complex, and is generated by the "summation of an exceedingly complex series of overlapping gradients of reinforcements, in part consisting of, but largely derived from, the 'primary' reinforcements occurring at the end of the temporal period covering the behavior sequence involved" (127, p. 143). In other words, more and more members of a stimulus series acquire (secondary) reinforcing power, each probably in accordance with the short gradient of reinforcement. The summation and interaction of these short gradients result in the extended *goal gradient*. Hull believes that the goal gradient is an exponential or negative growth function (not, as he had originally believed, a logarithmic function). The greater the influence of secondary reinforcement, the less steep the slope of the goal gradient, i.e., the greater the temporal range over which a reinforcement can exert its influence.

¹⁶ Skinner (237) found that delays are detrimental only after periodic reconditioning whereas delays up to 4 seconds did not affect the original conditioning of the lever-pressing response.

¹⁷ The expression *gradient of reinforcement* was first used by Miller and Miles (178).

Hull's goal gradient hypothesis (120, 122) has proved to be a powerful deductive tool by means of which a considerable amount of empirical results could be predicted, sometimes with striking accuracy. Among the findings which the hypothesis predicts and which were empirically confirmed are the following:

1. With number of reinforcements held constant, when the delay of reinforcement is short, less time is required to execute a response than under conditions of long delay of reinforcement (6).
2. Of a pair of alternative acts the one which is reinforced with a shorter delay is chosen (5, 226).
3. Other things being equal, the greater the difference in the delays of reinforcement yielded by two alternative reactions, the more quickly the animal will learn to choose the act yielding the shorter delay (5).
4. When the absolute differences in delay are equal, differentiation between two short delays is achieved faster than differentiation between two long delays. For example, a 30-second delay is more readily differentiated from a 60-second delay than a 60-second delay is from a 90-second delay (5, 85). On the basis of the goal-gradient hypothesis Hull has also been able to predict the effects of delays of equal relative, but of different absolute, duration (327).

The gradient of reinforcement in Hull's treatment is, strictly speaking, a *temporal* gradient. As Hilgard and Marquis (108) point out, such temporal gradients should be clearly distinguished from non-temporal gradients which refer to the spatial distance between a response and the reinforcement or the serial position of a response with respect to the reinforcement. Spatial separation or remoteness in a series, of course, implies temporal delay since it takes time to traverse the space leading to the goal or the members of a series ending in reinforcement. On the other hand, an animal delayed in a restraining compartment is in a very different situation from a subject that has to cross a runway in order to reach a food reward. The main difference lies in the fact that the spatial interval is filled with a series of acts leading to the reinforcement whereas sheer temporal delay may serve to disrupt the integration of a behavior sequence. The rat's speed-of-locomotion gradient (122) in a straight alley is a spatial gradient, with running speed plotted against successive segments of the path leading to the reward (6, 122) or punishment (23, 176). Similarly, the extensive experimental investigations of the goal gradient in maze learning (120, 238, and many others) and of animals' ability to discriminate short from long paths to reward (35, 50, 85, 327) should be considered primarily as studies of the *non-temporal* aspects of the gradient of reinforcement. In such experiments, the crucial experimental variable is the *distance* between a segment of the apparatus (runway, maze) and the locus of reinforcement. This distance involves not only a temporal delay but also a complex sequence of intervening acts.

A notable example of a spatial gradient is, of course, the Thorndike effect. The close kinship between the Thorndike effect and the goal-gradient was demonstrated by Muenzinger, Dove and Bernstone (198). In an "endless" maze (four identical mazes arranged as the sides of a square, with food boxes at the four corners) a double-winged gradient of elimination of errors was obtained. These authors believe that the goal-gradient is bidirectional in its fundamental form and that the usual backward elimination of errors in a maze reflects only the first half of the gradient. The other half cannot manifest itself because the animal's activity usually ends at the goal. This analysis has not remained undisputed. Hill (109) reports that the forward gradient appears only late in learning and ascribes it to the failure of anticipatory errors to be eliminated. In a more recent investigation, however, Thompson and Dove again report evidence for a basically bidirectional goal-gradient (262).

With human subjects, gradients of reinforcement have usually been plotted as a function of serial position rather than in terms of sheer temporal delay. The various studies of the Thorndike effect are a case in point: the spread of effect is most clearly demonstrated when frequency of repetition is plotted against distance from reward in terms of response units (15, 270, 331). On the other hand, when time alone is considered, a 6-second delay is found to be as beneficial to learning as a 0-second delay of reinforcement, and there is no evidence for a temporal gradient (167). The activity filling the interval between response and reinforcement is an important factor: when an announcement of *Right* follows an interval filled by another response, i.e., if the reward refers to the next to last response, the reinforcement is virtually ineffective. Response and reinforcement must "belong" together. (Such detrimental effects of interpolated activities are, of course, well known in the study of retroactive inhibition.) The fact that the gradient of reinforcement, which is readily demonstrated with animals, cannot be easily applied to human subjects is of considerable theoretical importance. With the aid of symbols the human learner can bridge temporal gaps that are prohibitive to the animal (192). Such apparent independence of immediate reward has considerably complicated the application of the law of effect to human learning.

Strength of Drive

The modern version of the law of effect equates effect to drive reduction. The strength of the drive could be expected to affect the operation of the law in two ways: (1) strength of the drive could be one of the determinants of the speed of acquisition and (2) performance (habit evocation) may vary with the strength of drive.

What is the influence of the strength of drive at the time of learning? An experimental investigation by Finan (73) showed that with the num-

ber of reinforcements constant an instrumental act is established more strongly when the drive is strong than when it is weak. A stronger drive yields a stronger habit although the relationship is by no means linear. This generalization has, however, not remained unchallenged. In Hull's theoretical system, the course of acquisition (building up of habit strength) is described as independent of the strength of drive at the time of learning. Habit strength is a joint function of number of reinforcements, the time for which the conditioned stimulus has been acting before the occurrence of the response to be learned, the time interval between response and reinforcement, and the magnitude of the goal object (127, 144). Drive strength is *not* one of the variables of which habit strength is a function. In a recent study Kendler (145) has justified this omission of drive strength as one of the determinants of habit strength. In one of his experiments, animals learned a bar-pressing response under different degrees of thirst deprivation. The animals were, however, equated in strength of motivation during extinction. The results show that different drive conditions at the time of learning had not influenced the amount of habit growth. Kendler also showed that a group of subjects which had a low drive strength during learning but received a large number of reinforcements established a stronger habit than a matched group which learned under high motivation but received a smaller number of reinforcements. The results reported by Finan and Kendler are contradictory and the role of the strength of drive during acquisition must remain open.

There is general agreement, on the other hand, that the strength of drive is an important determinant of performance. According to Hull reaction-evocation potentiality is a multiplicative function of habit strength and drive strength. This hypothesis receives experimental support from Perin's finding that for a given number of reinforcements, resistance to experimental extinction is an almost linear function of the number of hours of food deprivation at the time of the extinction procedure (209). In his investigation of operant conditioning, Skinner (237) found that rate of response is considerably affected by the animal's drive: over a wide temporal range of food deprivation, increase in drive leads to faster rate of emission of responses (emptying of the reflex reserve).

Although performance varies with strength of drive, there are some indications that appropriate stimuli will evoke a response even when the subject is to all intents and purposes fully satiated (209, 330). Extrapolation from Perin's theoretical curves relating resistance to extinction to drive strength predict this result. But here, too, the last word has not been spoken. Koch and Daniel (147) report zero or close to zero reaction potential immediately after satiation. Although the general dependence of performance on strength of drive may be considered as

well established, the precise quantitative nature of this relationship as well as the problem of interaction of different drives (144, 145) is still in need of further experimental analysis.

THE PLACE OF THE LAW OF EFFECT IN LEARNING THEORY

Parametric studies such as those described in the preceding section derive their main significance from the contribution which they can make to general learning theory. We shall now consider the role played by the law of effect in theoretical interpretations of learning.

Theories of learning can be classified in more than one way. One can distinguish between molar and molecular theories, according to emphasis on specific movements as against stress on acts and their outcomes; or one can pit modern associationism against configurational theories. The role assigned to reinforcement by reward and punishment provides another criterion of classification. There are clear-cut lines of division separating systematic points of view according to the role assigned to the law of effect: (1) effect may be considered to be a principal condition of all learning, (2) effect may be rejected as a condition of learning but considered an important determinant of performance, (3) the essentials of the learning process may be conceptualized without reference to effect, with reward and punishment assigned a subsidiary role and credited with only indirect influence on learning.

The Law of Effect as a Principal Condition of all Learning

The law of effect could not become the pivot of a comprehensive theory as long as it was restricted to the narrow universe of multiple-choice ("trial-and-error") learning. It was only with its application to the facts of conditioning that the law of effect could become a unifying principle around which a systematic theoretical structure could be built.

For a considerable period of time students of effect and students of conditioning failed to make contact with each other. In the field of classical conditioning stimulus *substitution* was the primary principle of explanation. To the extent that substitution could not account for all the empirical data, auxiliary hypotheses such as the principle of dominance and such concepts as set and attitude were introduced (105, 106). On the other hand, Thorndike firmly maintained the distinction between learning by selection and fixation (formation of S-R bonds) and conditioning, which he termed "associative shifting." Thorndike never believed that the study of conditioning could throw much light on the nature of learning: "The conditioned reflex is one type of learning that manages, even more completely than maze learning, to conceal the true nature of the learning process in a mass of special conditions" (267, p. 85). In the light of this opinion it is not surprising to find that Thorn-

dike has almost completely ignored the facts and theories of conditioning in all his writings. Students of conditioning, on the other hand, could not indefinitely ignore the problem of effect, for the crucial role of incentive in the establishment and maintenance of conditioned responses was an incontrovertible experimental datum.

Paralleling Thorndike's dichotomy between connection formation and associative shifting, several two-fold classifications of learning situations have been made: classical and instrumental conditioning (108); Type S and Type R conditioning (236); conditioning and success learning (228); quantitative and qualitative conditioning (216); conditioning and motivated learning (74). Such classifications reflect important differences in experimental procedures under which learning can take place, but, as Hilgard and Marquis (108) have emphasized, they do not represent pure types of learning which necessarily require principles of explanation as different as substitution and effect. In any given experiment, both types of learning may take place, and the classification of an experiment will largely depend on which aspects of the response the experimenter emphasizes and measures. A classical conditioning experiment emphasizes stimulus substitution and homogeneous reinforcement; an instrumental reward or escape experiment dramatizes the principles of effect and heterogeneous reinforcement. "It is a common error to permit the reference experiment to dramatize a particular process, and then to suppose that the experiment represents a pure case of the process dramatized" (108, p. 97).

Thus, the distinction between learning by selection and fixation (heterogeneous reinforcement) on the one hand and learning by stimulus substitution (homogeneous reinforcement) on the other does not preclude a unified conceptual scheme which makes a law of effect the basic principle of all learning. It is the virtue of Hull's theoretical analysis that these two types of learning are subsumed under a common principle: they are both reduced to the operation of the law of primary reinforcement (Hull's formulation of the law of effect). The basic principles and generalizations of Thorndike and Pavlov are combined and unified in a single theoretical structure (246). Hull's law of primary reinforcement makes temporal proximity to a reinforcing state of affairs a condition without which learning cannot take place. Both selective learning (the Thorndike situation) and conditioned-response learning are special cases of the operation of this law. In the case of simple selective learning, one of many possible alternative reactions occurs in temporal proximity to need reduction and hence is differentially reinforced. Such a receptor-effector connection may or may not be of super-threshold strength at the beginning of an experiment. The conditioned response also depends on temporal proximity to reinforcement but in this case a *new* receptor-effector connection is almost invariably established: the connection between CS and UR. Hull concludes that "the differences between the

two forms of learning are superficial in nature, i.e., that they do not involve the action of fundamentally different principles but only the differences in the *conditions* under which the principle operates" (127, p. 78). When the CR is considered basically akin to selective learning, the behavioral laws discovered in conditioning experiments can be applied to, and integrated with, the phenomena of selective learning. The principles of conditioning in conjunction with the law of primary reinforcement can be used in the deduction of more complex forms of learning.

Although the law of primary reinforcement is closely akin to Thorndike's law of effect, there is an important difference which Hull has made explicit. As the law has been repeatedly stated by Thorndike it is both a law of motivation and a law of learning. In a review of Thorndike's work Hull (123) raised the question whether motivation (striving) produces the learning, or learning produces the motivation, or whether some third and still more basic process produces both. Thorndike's formulation seems to imply that striving is to be considered primary: he defines a satisfier as that which an animal strives to attain or does nothing to avoid. Hull, on the other hand, considers learning ("strengthening") primary and derives striving from the principles of conditioning and the primary law of reinforcement (need reduction) as basic assumptions. In his own words, "states of affairs which organisms will strive to attain are reinforcing agents, not because they will evoke striving, but they evoke striving now because at some time in the past they were potent reinforcing agents, thereby joining stimuli and responses . . . which constitute the striving" (123, p. 822). The point is that Hull is willing to assume as originally given only those reinforcing agents which satisfy basic biological needs and are linked with the organism's survival, and he then derives other motives or strivings with the aid of the principles of conditioning. In Thorndike's writings, on the other hand, we find no such hierarchy of motives. The role of motives (wants, interests and attitudes in learning) is described as two-fold: (1) they determine what response a situation shall evoke, and (2) the satisfaction of wants strengthens S-R bonds (276). It is to the question of the *origin* of these wants, interests and attitudes that Hull addresses himself, seeking to derive them with the aid of conditioning principles.

The gap between learning situations in which biological need reduction occurs and the myriad of learning situations in which there is no such immediate primary reinforcement is bridged by the principle of *secondary reinforcement*. According to this principle, any receptor activity which regularly precedes a primary reinforcement will itself gradually become a reinforcing agent. The sight and smell of food as well as other stimuli emanating from a feeding compartment, such as the click of the mechanism releasing pellets into a Skinner box (22, 237),

are sources of secondary reinforcement. A stimulus such as a buzzer or tone which has been regularly associated with shock also acquires reinforcing properties (67, 69, 74). The reinforcing power of token rewards, i.e., rewards which may subsequently be exchanged for a primary reward such as food may similarly be explained in terms of derived reinforcement (39, 65, 179, 207, 319). The effectiveness of social rewards and punishment calls for similar explanatory concepts: a practically unlimited chain of higher-order conditioning must be assumed (127, 177).

Recently Spence (247) has suggested that it is the particular stimulus *pattern* at the time of reward which acquires secondary reinforcing properties. This reinforcing power is generalized to preceding stimulus patterns according to a temporal gradient. In this conceptualization the vexatious problem of backward action is eliminated.

It is in Hull's theoretical system that the law of effect has come to its full theoretical fruition: it has become a primary postulate with whose correctness a complex structure of deductions and theoretical interpretations must stand or fall.

The Law of Effect as a Law of Performance

It is possible to accept the law of effect in a *descriptive* sense, i.e., to recognize that rewarded responses are usually repeated in preference to non-rewarded ones, and yet to deny that the law of effect is a law of *learning*. Several writers, under the leadership of Tolman, have taken this position (64, 159, 287, 289, 292, 294, 315). A rigorous distinction is made between the *acquisition* and *utilization* of habits. The acquisition of habits depends on the formation of *cognitive* patterns within the organism which reflect the stimulus relationships in the environment. Such cognitive patterns Tolman has described as "sign-gestalt expectations" or "hypotheses" (287, 288, 290, 291). Organisms come to accept one event as a sign or "local representation" of another event (293). Learning occurs when the subject has built up an expectation that a given sign in the environment will, via a behavior route, lead to a certain significate or outcome. These expectations result from the organism's commerce with the environment and their acquisition is governed by such conditions as frequency, recency, and perceptual laws of stimulus organization (sign-gestalt formation).¹⁸ Differential reward is not considered a determinant of learning. Experienced reward does, however, play a role as a determinant of performance or utilization of habits. From a set of alternative responses to a stimulus (sign) that response is *selected* and *performed* whose consequence is most "demanded," i.e., *most rewarding in terms of the momentary motivational state of the*

¹⁸ In building up expectations the organism essentially reacts to the relative probabilities that signs in the environment will be followed by certain outcomes. A reinforcement theory stressing reaction to probability has been elaborated by Brunswik (21).

animal. Even though knowledge and need (learning and performance) are thus distinguished, behavior is always a joint function of both. When a subject moves toward a goal he must (1) have a need for that particular goal, and (2) have the knowledge that a given piece of behavior on his part will lead to that goal. Hence needs and knowledge constitute an interdependent pattern or field (315). In this theoretical account, then, the law of effect is rejected as a law of *learning* and relegated to a secondary role as a condition of the moment-to-moment utilization of habits that have been acquired independently of effect. Only if motivation remains constant can behavior be accurately predicted on the basis of the law of effect.

Tolman calls his account of learning and performance a field theory¹⁹ which may be applied to substitute stimulus (CR) learning, trial-and-error as well as more complex forms of learning such as "inferential" and "inventive learning" (290). The learning theory put forward by the chief exponent of field theory, Kurt Lewin, (161) is closely akin to that of Tolman. Lewin sharply distinguishes between learning as *change in knowledge* or cognitive structure (differentiation of unstructured areas, restructurization) and learning as *change in motivation* (changes in valences and values). Changes in cognitive structure are ascribed in part to the same type of "forces" as govern perceptual fields, in part to the impact of the needs, values, and hopes of the subject. Among the complex of forces governing changes in motivation reward is only a minor factor, for "forces governing this type of learning are related to the total area of factors which determine motivation and personality development" (161, p. 239). Throughout his treatment Lewin emphasizes the need to distinguish the motivational from the cognitive problems, to study their separate laws, and then to determine the role of each type of factor in different learning situations. In this connection it should be noted that the distinction between learning and performance is not limited to the field-theoretical approaches. In Hull's account, a parallel distinction is made between the principles governing habit-formation and the principles governing habit use. Whereas the concept of *habit strength* describes the degree of acquisition, the construct of *effective reaction potential* refers to the degree to which a habit is ready for performance.²⁰ Thus both S-R theory and field theory allow for the distinction between learning and performance. The difference is that for Hull motivational factors govern both acquisition and performance, whereas in the Tolman-Lewin formulation motivational factors come into play only in the utilization of habits.

One of the main experimental supports for the Tolman-Lewin view

¹⁹ For another restatement of learning theory in field-theoretical terms, see Adams (1).

²⁰ Effective reaction potential is a product of effective habit strength and drive, taking account of the total amount of inhibitory potential (127).

has been the phenomenon of latent learning. When a reward is introduced after a series of unrewarded trials in a maze, an improvement in performance occurs which far exceeds the usual effects of a single reward. A substantial part of this sudden improvement may then be ascribed to learning (formation of sign-gestalt expectations) which took place during the unrewarded trials but was not utilized in the absence of reward. Such analysis is strictly in accord with Tolman's theory. The phenomenon of latent learning was first demonstrated by Blodgett in 1929 (14) and has been repeatedly confirmed since (49, 51, 102, 103, 295, 304). A recent repetition of Blodgett's experiment by Reynolds (220), however, failed to show latent learning. This result has, at least temporarily, detracted somewhat from the support which the latent learning experiment has provided for the field-theoretical view. To the extent, however, that latent learning has been successfully demonstrated it provides a serious challenge to any view that would make learning a cumulative function of successive reinforcements by reward. As Leeper (160) has pointed out, it is impossible to argue that some other reward, such as satisfaction of the exploratory drive, accounts for the learning. Such a reward should reinforce exploratory behavior and not running to the goal with few errors once a food reward is introduced. In this connection it is interesting to note that in a recent attempt to derive the facts of latent learning within the framework of Hull's theoretical system Seward (232) was forced to abandon Hull's postulate (No. 4) that increments from successive reinforcement summate to yield a combined habit strength. Instead, he had to assume that conditioning is independent of reinforcement and is complete in one trial when S and R are simultaneous. On the other hand, Buxton (28) was able formally to derive the phenomenon of latent learning in terms of field-theoretical principles.

Utilization of a habit depends on the presence of appropriate motivation (appetite for the goal object). Performance of a given act does not always depend on the presence of any *one* particular motive: different motives may be equivalent to each other in their capacity to evoke performance to the extent that the outcome remains congruent with the motive that prevails at the time of performance (61, 62, 63, 64). Within limits, therefore, changes in motivation may leave behavioral responses (habit utilization) relatively unaffected, or lead to only partial changes in performance. Even total removal of reward may lead to an only temporary disturbance in performance rather than a permanent disintegration of a habit (310). When the subject is confronted with a *choice*, however, motivation may serve as the basis on which *selection among alternative responses is made*. On this hypothesis, if a subject were to learn that one route in a maze leads to food and another route leads to water, he would be expected to choose the food route when hungry and the water route when thirsty. This deduction has been repeatedly put

to experimental test but the results and interpretations have thus far remained contradictory.

In 1933 Hull (121) performed an experiment in which he was able to establish differential reactions to the same maze environment on the basis of different drives (hunger and thirst). He trained his rats in a simple two-route maze. The animals had to traverse one arm of the maze when they were hungry and were fed in the goal chamber. On days when they were thirsty, the animals had to traverse the other arm of the maze and received water in the *same goal chamber*. Hull's animals learned this discrimination only with great difficulty: 25 training periods of eight days each were needed before 80 percent correct responses were given on the first run of a series. Hull interpreted his results to show that animals can be conditioned to internal conditions (drive stimuli). The slowness and difficulty of the training would seem to run counter to the predictions of a sign-gestalt theory. The experiment was then repeated by Leeper (159) who fully realized its crucial significance for a field theory of learning. He introduced an important modification. In order to make a clear perceptual differentiation of the situation possible for the subjects, he constructed a maze with *two end-boxes*, one containing food and the other water. When a rat made an incorrect choice, it entered the goal chamber which contained the reward not desired under its prevailing motivation. Thus the rat had continual opportunity to build up an expectation of "what leads to what." Whereas in Hull's experiment the rat was simply blocked on an incorrect trial, Leeper's subject acquired information about means-end relations on every trial, correct and incorrect. As a result Leeper's rats required only one eight-day period to reach approximately the same criterion as Hull's rats after 25 eight-day periods.²¹ Leeper felt that his results fully support Tolman's theory and the distinction between acquisition and utilization of habits.

The results of Hull's and Leeper's experiment are reviewed in some detail to show how virtually the same experiment lends itself to alternative interpretations, one based on the law of effect and assuming summation of reinforcement, the other making learning independent of effect. Subsequent studies of the same problem have remained equally inconclusive. Kendler (146) raised the question whether results similar to Hull's and Leeper's could be obtained if both hunger and thirst drives were simultaneously present during the training period. He found that animals learned to respond appropriately on the test trials, i.e., in accordance with the motivation prevailing during the test trials. Although this result would seem to be exactly in accord with Leeper's

²¹ Hull explained the difference between his and Leeper's results by suggesting that after the first few trials Leeper's animals probably operated under both drives, thus being rewarded no matter which compartment they entered (127). In a reply to Hull, Leeper denied on the basis of his experimental records that such was the case (160).

interpretation, Kendler offers alternative explanations within the framework of an effect theory. He speculates that (1) only those drive stimuli which are reduced during the training trial are associated with the rewarded response, and (2) invokes anticipatory goal reactions as possible differential cues.

Recently Spence and Lippitt (249) have reported "an experimental test of the sign-gestalt theory of trial and error" which is again concerned with the utilization of alternative habits under different motivations. The subjects (white rats) in this experiment were motivated by thirst and given 12 days experience in a simple Y-maze. One arm of the maze always led to water. The other arm led to food for one-half of the subjects, to an empty goal box for the other half. The test trials were run under hunger motivation, with the thirst drive satiated. During these test trials the subjects continued to go down the alley leading to water and the group which had previously experienced food was not superior to the other group in learning to choose the food alley. These results are clearly contrary to what would be expected in terms of sign-gestalt theory. However, the same authors had previously reported (in abstract form) an experiment which was more in accord with Tolman's theory (248). In this case, the subjects were satiated for both food and water during the training series and found water at one end of the maze and food at the other. When made hungry or thirsty, the animals chose the alley leading to that goal which satisfied the need prevalent at the moment. At that time Spence and Lippitt concluded, "Latent learning does not occur in the situation where animals perceive the subsequent goal object while motivated for another, but latent learning does occur where complete satiation made for no particular goal-directedness." Thus the results of Spence and Lippitt do not call necessarily for abandonment of a sign-gestalt interpretation but for a modification: experience with goal objects which satisfy a need and with goal objects for which there is no need during the training period are not equally effective in learning (acquisition of expectations).

Just as the facts and interpretations of latent learning remain under discussion, the general problem of the mechanism of discrimination learning is still the cause of serious disagreement between proponents and opponents of an S-R-effect theory of learning. The experimental fact, on which there is general agreement, is that subjects can learn to choose between positive stimuli and negative stimuli simultaneously presented, i.e., between stimuli response to which leads to reward and stimuli response to which fails to lead to reward or leads to punishment (150, 154, 156, 157, 170, 205, 239, 241, 245). Even before the discrimination has been learned, the animal does not respond in a haphazard fashion but shows definite systematic response tendencies, or "hypotheses" (149, 150, 151, 152). It is around these systematic response tendencies during the pre-solution period that the main argument be-

tween "continuity" and "non-continuity" theories has been centered. According to the continuity theory, discrimination learning, like other types of learning, results from a cumulative process of building up an association between the positive stimulus cue and the response. Every time a response to the positive stimulus cue occurs and is followed by reward, the association between cue and response is strengthened; every time a response to the negative stimulus cue is made and fails to be followed by reward, the tendency to respond to this cue is weakened. Discrimination is established when the difference in the excitatory strengths of the positive and negative cues is sufficiently great to overcome other aspects of the total stimulus situation which are not consistently associated with reward or failure. In terms of this analysis discrimination learning is fully explained in terms of stimulus-response association and the law of effect. The continuity hypothesis of discrimination learning has been sponsored by Spence (240, 241, 242, 243, 244, 245), and McCulloch (168, 169, 170). Spence has shown that a derivation of "hypothesis" behavior during the pre-solution period is possible in the framework of the continuity theory (240) as is a derivation of the special type of discrimination learning studied in "transposition" experiments (242). According to the noncontinuity theory as it is interpreted today, the animal learning a discrimination changes from one systematic mode of response (hypothesis) to another until the problem is solved but *practice on the unsuccessful hypothesis does not contribute to the learning of the correct association*. This view has been defended primarily by Lashley (157, 158), Kreshvesky (153, 154), and Haire (97, 98, 99).

Both theories have been supported by important experimental investigations. A survey of the evidence suggests that the argument is still in an inconclusive stage: The crucial experimental question is whether associations are formed during the *pre-solution period* which significantly influence the subsequent establishment of the discrimination. In a typical investigation of this problem, the significance of the stimulus cues is reversed during the pre-solution period; the cue which is to be positive on the test trials is made negative and vice versa. According to the continuity theory such reversal should slow down learning on the test trials, according to the non-continuity theory the reversal should have no effect. The first demonstration of the cumulative effect of training was presented by McCulloch and Pratt (170) in their study of weight discrimination by white rats. Preliminary training to the lighter of two weights slowed up learning when the heavier weight was made positive. In a study of visual form discrimination habits of chimpanzees Spence (241) similarly found that the establishment of a discrimination was directly dependent on the excitatory strengths of the positive and negative stimuli. The greater the relative number of reinforcements a given stimulus had received in a

series of learning tasks, the easier it was to establish a positive response to that stimulus. The correlations obtained were positive, high and significant. Spence also showed that sudden learning (insight) like gradual learning, was positively correlated with the excitatory strengths of the positive and negative stimuli as determined by the number of reinforcements (243).

Not all the evidence has been on the side of the continuity theory, however. Using the technique of reversing the positive and negative stimuli during the pre-solution period, Kreshevsky (154) found that the reversal did not significantly affect the learning of the correct solution. Lashley (157) after repeating Spence's study of form discrimination with rats argued that evidence from positive correlations between number of reinforcements and ease of discrimination learning is inconclusive. He considers the assumption unwarranted that a high correlation between two arrays cannot exist when the values of one are in part determined by chance. In computing his own correlation Lashley omitted all trials showing systematic reaction to position, thereby changing the excitatory values of the stimuli by random measures from 0 to more than 100 percent, and yet the correlations between number of reinforcements and ease of discrimination learning were not reduced but slightly increased! Lashley then offered experimental proof against the assumption that all stimuli acting at the time of a response tend to be associated with that response. Rats who learned to respond in terms of size did not at the same time learn to respond in terms of that shape which was consistently associated with the positive stimulus. "If the animals are given a set to react to one aspect of a stimulus situation . . . large amounts of training do not establish association with other aspects, so long as the original set remains effective for reaching the food" (157, p. 259). Instead, Lashley emphasizes the role of perceptual organization and attention in determining which aspects of the stimulus situation will be associated with the response.

The controversy still continues. In 1945 Spence (245) published a carefully designed experiment in which he again used the technique of cue reversal. Great care was taken to control all relevant factors including position habit. Cue reversal significantly retarded learning and again led Spence to conclude that the development of an association between a cue and a response is a cumulative process which is independent of systematic response tendencies to other cues during the training period. Discrimination depends on the number of times the animal has been rewarded in the presence of the positive cue.

In spite of Spence's impressive results, the last word has not been spoken. Recently Lashley and Wade (158) published a stringent criticism of "neo-Pavlovianism." They singled out for theoretical and experimental attack two assumptions central to "neo-Pavlovianism," i.e., to modern effect theory: (1) that in conditioning all aspects of the

stimulus situation are associated with the reaction, and (2) that the effects of reinforcement spread to stimuli other than those present during training (stimulus generalization). Lashley and Wade believe these principles to be contrary to fact. In their experiments, groups of subjects (rats, monkeys) were trained to react positively to a given stimulus. The stimulus would then be opposed in a discrimination experiment to another on the same stimulus dimension. In some cases, the reaction to the initial stimulus was reinforced, i.e., the initial stimulus remained positive; in other cases, the reaction to the initially positive stimulus was extinguished, i.e., it was made negative in the differential training. The rates at which discriminations could be established under these two conditions were compared. In every case differential training was faster when the initial reaction was extinguished than when it was reinforced! The differences were consistent though not statistically reliable. On the basis of these results Lashley and Wade reassert that in discrimination relatively few, often not more than one, aspects of the stimulus are effective in the choice reaction of the animal. "A definite attribute of the stimulus is 'abstracted' and forms the basis of reaction; other attributes are not sensed at all or disregarded" (158, p. 81). Stimulus generalization is ascribed to failure of association. When a subject responds to a stimulus to which he was not originally trained, he does so because of his failure to attend to those characteristics which distinguish the training stimulus from the stimulus to which the reaction is generalized. On the other hand, a subject establishes a differentiation when he forms associations with such aspects of the stimuli as he had not attended to in earlier stages of his training. The burden of the argument is a denial of automatic irradiation and indiscriminate association of stimuli with responses by virtue of sheer temporal proximity to a reinforcing state of affairs.

The crucial conceptual distinction which the field-theoretical approach has introduced is between learning and performance. Though recognizing the logical status of this distinction McGeoch seriously questioned its experimental usefulness. "The only way we can know that learning has occurred is by an observation of successive performances since *learning is a relation between successive performances*. . . . Assertions that motive and effect influence performance but not learning become meaningless in the absence of quantitative demonstration, a demonstration which cannot be made without measurements of performance" (171, p. 599). Replying to this criticism, Leeper (160) pointed out that different test conditions may result in very different performances following the same learning situation. In strict conformance with McGeoch's view one would then have to conclude that any learning situation results in a large number of "learnings." A much more parsimonious approach consists in the use of such distinct intervening variables as learning and performance, or, in Hull's language, habit-formation and habit-evocation.

Another difficulty which results from the distinction between learning and performance is the conceptual gap which is left between knowledge and action. If learning consists in the building-up of expectations, how is expectation translated into action, and how can the specific character of action be predicted? The field theory predicts what a subject will come to expect but it fails to predict in any specific way what he will do as a result of the expectation. Guthrie, who has long been concerned with the analysis and prediction of particular movements, has expressed this criticism as follows: "Signs, in Tolman's theory, occasion in the rat *realization*, or *cognition*, or *judgment*, or *hypotheses*, or *abstraction*, but they do not occasion action" (90, p. 172). Hilgard and Marquis (108) regard this failure to predict specific action as a weakness but also see in it a source of strength. A breadth of interpretation is possible which strict conditioning theories do not allow. A variety of performances can be grouped together in terms of the purpose which they serve without regard to particular details of movement.

Rejection of the Law of Effect as a Basic Principle of Learning or Performance

It is his insistence on an analysis of the learning process in terms of specific stimuli and specific response movements which has led Guthrie (88, 89, 90, 91, 92, 93, 94) to the formulation of a theory in which the law of effect has no place as either a basic principle of learning or of performance. Guthrie distinguishes strictly between acts and movements. An act is a class of movements defined by the *end result*. The law of effect with its emphasis on the *consequences* of S-R connections thus applies primarily to acts. In Guthrie's view, however, it is to movements and not acts that the basic laws of learning must refer. The achievement of an act, i.e., of an effect or end result, is "completely dependent on the acquisition, through learning, of a specific stereotyped movement or set of movements for the accomplishment of the effect that defines the act" (94, p. 51). As to the association of stimuli and response movements, it is entirely explained in terms of one basic principle—association by contiguity. It is always the last movement or set of movements made *simultaneously* with a given stimulus that is repeated on recurrence of the stimulus. Only a single coincidence of stimulus and response is necessary to establish the association. Thus learning is accomplished in one trial but so is unlearning, for one presentation of the stimulus on which the response is not made destroys the association. Although all learning is complete in one trial, the process of acquisition proceeds only gradually because of the continuously changing nature of the components of the stimulus situation. Only when the response movements have been conditioned to the various components of the stimulus situation—exteroceptive, proprioceptive, and interoceptive—can the response be reliably evoked.

Guthrie's description of the basic processes in learning is made independently of the action of rewards and punishments. The practical efficacy of reward and punishment can be satisfactorily accounted for in terms of stimulus-response contiguity. It is always the response last associated with a stimulus which will be repeated when the stimulus recurs. A successful response either removes the organism from the stimulus situation (for example, by allowing escape from a puzzle-box) or so alters it (for example by removal of internal drive-stimuli through feeding) that no new associations with the situation can be formed. Similarly the effectiveness of punishment depends on what it makes the organism *do* in the presence of a stimulus (for example, make withdrawal movements), not on what it makes him feel. It is the function of reinforcing agents to protect the associations made. Effect prevents unlearning and allows the law of recency to operate.

Guthrie's explanation of effect in terms of changed stimulus context has been criticized on the ground that rewards may be effective even though they do not clearly alter the stimulus situation (169). Eating a small pellet of food, for example, does not remove the drive stimulus and yet is effective in strengthening an S-R connection. Guthrie believes that even with a small reward the stimulus situation is materially altered since the "annoyance"—the restless, excited behavior which precedes feeding—is removed though the "annoyer" (the drive stimuli maintaining the animal's search for food) persists. Removal of annoyance removes the cues for successful action and thus prevents unlearning (92). This reformulation is not entirely conclusive. After eating, the animal may soon again be in an aroused state, the annoyance may return, and previous responses would be expected to be unlearned. "It is obvious that something has changed when the animal ceases to run or to push levers and begins to eat, but a theory must state more precisely just what the change is which guarantees the learning of the prior response if the theory is to be verified experimentally" (108, p. 92).

As to experimental verification of the theory, the most outstanding work is that of Guthrie and Horton (95) who reported a detailed, well documented study of cats' behavior in a puzzle box. The filmed records of the animals learning to escape from a puzzle box are well in accord with Guthrie's theoretical expectations. The most striking characteristic of the subjects' behavior was its repetitiousness. The movements immediately preceding escape were distinguished by their stereotypy since removal from the situation presumably protected the last association formed. Whatever variability of behavior appeared is explained by changes in the stimulus situations which caused new associations to be established.

More recently, an experimental study based on Guthrie's theory was reported by Seward (230) who was interested in testing the "finality theory of reinforcement." He divided his subjects (rats) into two

groups. One group was given food upon pressing a bar, the other group was removed from the experimental situation when it had made the response. Though both groups learned to press the bar, the reward group was clearly superior to the removal group. Though removal from the situation protects the last association, the total effectiveness of reward cannot be ascribed to the termination of the situation.

Stimulus-response contiguity theory has been criticized as not easily lending itself to experimental test (108, 246). The continually changing stimulus elements, especially the subjects' own proprioceptive impulses, are not readily controlled and manipulated. Usually changes in the stimulus situation have to be inferred from the failure of a response to be repeated. Such inferences are, of course, circular. The main difficulty lies in the highly specific nature of the stimuli and responses in terms of which the analysis is made. The theory is distinguished by logical and consistent formulation; it exemplifies a coherent account of learning developed without recourse to a law of effect. But it is still greatly in need of experimental verification.

THE ROLE OF EFFECT IN COMPLEX LEARNING

Our survey has shown the pivotal role of the principle of effect in the construction of a systematic theory of learning. Experimental tests of the various theoretical propositions have been conducted almost exclusively in the animal laboratory (Hull, Tolman, Guthrie) or with human subjects in simple rote-learning situations (Thorndike). "Proofs" and "disproofs" of the law of effect in such experiments still leave open the question of what role effect (satisfaction) plays in more complex types of learning. The question is sharpened by two types of observation: on the one hand much learning occurs in the absence of demonstrable drive reduction, and on the other hand satisfying responses sometimes fail to be repeated (3). The challenge of complex adult learning which so often seems to defy the principle of effect, can be met in two ways. It is possible, with Mowrer (185), to defend the law of effect as a universal principle of learning despite the apparent failure of many learners to obey it—by reformulating the law of effect so as to encompass these more complex forms of learning. On the other hand, one may assert, as Allport has done (3, 4), that the law of effect holds only for animals, small children, mental defectives and in some peripheral phases of adult learning, whereas different principles, such as ego-involvement and active participation, need to be invoked in the analysis of more complex adult learning. Thus, for Allport the law of effect is not a *law* of learning but merely one of the many conditions which may favor learning—a condition, moreover, which applies only to a very circumscribed segment of learning behavior. The controversy between

these two views is still in progress and is well exemplified by the recent symposium on "The Ego and the Law of Effect."²²

In attempting to extend the law of effect to learning which is not motivated by physiological need reduction the concept of *derived* or *secondary* drive plays a central role. A secondary drive is a learned drive. "These secondary drives are acquired on the basis of primary drives, represent elaborations of them, and serve as a facade behind which the functions of the underlying innate drives are hidden" (177, p. 19). Under the heading of derived drives come anxiety and anger and such social needs as pride, ambition, and desire for social approval. Miller and Dollard (177) suggest that self-induced stimulation is the basis of acquired drives as well as the basis of acquired rewards and purposes. "Most of the responses which are the basis of socially significant acquired drives and acquired rewards are internal responses . . ." (p. 55). In the course of learning, any stimulus cue which acquires the ability to reduce a secondary drive acquires reward value.

Secondary drives are learned but they may also serve as the basis for further learning since the reduction of a secondary drive is a reward which may reinforce S-R connections in accordance with the law of effect just as does the reduction of a physiological drive. By postulation of such derived drives the gap between animal learning and social learning may be bridged and analysis of the latter in terms of stimulus, response, and effect (tension reduction) is made possible. Miller and Dollard's book, *Social Learning and Imitation* (177) represents an attempt to apply this analysis to such social behavior as imitation, leadership and crowd action. May (175) has attacked the problem of war and peace in a similar conceptual framework, and Whiting (316) has applied stimulus-response-effect analysis to the process of socialization in a primitive society. In terms of such analyses, a complex network of acquired drives develops out of the primary biological drives, and it is through the reduction of these acquired drives that the process of socialization operates. When adult learning *seems* to proceed without obvious drive reduction we must search for the derived drives whose satisfaction makes learning possible.

The concept of derived drive has been experimentally fruitful. A series of experiments conducted over the last ten years or so by Mowrer is a notable demonstration of the extent to which this concept has produced testable hypotheses. Mowrer has made a determined effort to explore the implications of the law of effect to their limits and to erect upon the foundations of this law a unitary account of learning which

²² *Psychol. Rev.*, 1946, 53, No. 6 (4, 185, 221).

would have equal applicability to animal learning and to complex social learning. The concept of derived drive appears to be the touchstone of this theoretical edifice. Mowrer has demonstrated, for example, that *preparatory set* or *expectancy* may function as a drive whose reduction serves as a motivating factor in learning (180, 183, 190). A state of tension or discomfort arises not only from the presence of a basic organic need but also from the anticipation of the recurrence of one or more of these needs. Not only does the *application* of an electric shock result in tension but also the *anticipation* of the shock, and this anticipation constitutes a derived drive. This state of tension is high before the occurrence of the punishment and low immediately thereafter (183). Reduction of this anticipatory tension is rewarding. "Other things being equal, the greater the extent of the drop in the expectancy-tension after the occurrence of a stimulus-response sequence the greater the reinforcing or learning-inducing value of this drop" (183, p. 38).

Mowrer's stimulus-response analysis of *anxiety* (181, 182, 190) and *fear* (189) is closely related to his conception of expectancy. Anxiety is the (learned) anticipation of a noxious stimulus. This anticipation, which is a source of tension and discomfort and hence acquires drive quality, results in a variety of acts from which are selected and fixated (by the law of effect) those forms of behavior which are most instrumental in the reduction of anxiety. The postulation of anxiety and anxiety-reduction explains the seemingly paradoxical finding that a conditioning procedure which permits *avoidance* of a shock results in better conditioning than a procedure which merely allows escape from the shock or mitigation of the shock (17, 190). Avoidance results in anxiety-reduction which provides reinforcement even though the noxious stimulus itself is not delivered. The effectiveness of avoidance and concomitant anxiety reduction depends in part on the temporal sequence of the noxious stimuli. If shocks are presented at regular temporal intervals, better conditioning results than with irregularly spaced presentation. When stimuli come in a regular order, anxiety mounts to a maximum as the time approaches for presentation of the stimulus, then drops. With irregular presentation, the subjects are "kept in a more or less chronic state of apprehension or suspense" (182, p. 510). Mowrer and Lamoreaux have also been able to show that in avoidance conditioning, the CS becomes a source of anxiety whose termination is rewarding in itself independently of the reinforcement provided by avoidance of the shock (189). As a result, conditioning is better if the CS (source of anxiety) terminates at the moment when the conditioned response occurs than if it ceases either before or after the response. The

fact that the CS in avoidance training becomes itself a source of anxiety and thereby a motivating factor in its own right helps to establish conditioned avoidance responses which are radically different from the responses made to the unconditioned stimulus. Thus some of Mowrer and Lamoreaux's rats were trained to jump in order to terminate the CS whereas the "correct" response to the UcS was running, and vice versa. Mowrer and Lamoreaux conclude that there are two distinct sources of reinforcement in avoidance conditioning, the one tending to strengthen the connection between fear and whatever response reduces the fear and the other which strengthens the connection between fear and whatever response eliminates the situation (shock and fear combined). They believe that the recognition of fear as a secondary motive in learning has important systematic implications for the law of effect. "In this way so-called anticipatory or 'foresightful' behavior can be made to conform very acceptably to the requirements of the Law of Effect, and seemingly 'purposive' or 'teleological,' responses are accounted for well within the framework of scientific causation" (189, p. 48).

Even though such intervening variables (secondary drives) as anticipation, anxiety, and fear²³ help to subsume under the law of effect behavior which is seemingly foresightful or purposive, a vexatious problem remains. Why do organisms so frequently persist in behavior which is clearly punishing, or, at least, more punishing than rewarding? It is, of course, possible to argue that whatever behavior subjects persist in must somehow be rewarding or it would be abandoned. Such reasoning would, however, be clearly circular and would beg the very question at issue in discussions of the law of effect. Recently Mowrer and Ullman (192) have attacked this problem of "non-integrative" learning. They believe that the key to the riddle of non-integrative learning lies in the temporal pattern of rewards and punishments. Many acts have consequences which are both rewarding and punishing. According to the gradient of reinforcement, those consequences which immediately follow the act will be more influential in learning than temporally remote effects. Integrative learning occurs when remote consequences are through symbolic behavior brought into the psychological present. Non-integrative learning takes place when the organism fails to make the symbolic bridge to the future and remains at the mercy of the im-

²³ Another derived drive is *fatigue*. In extinction, fatigue is the motivation competing with the performance of the act. Not to respond reduces the fatigue and is therefore rewarding. This analysis is supported experimentally by the fact that effortfulness of task is inversely related to the number of extinction responses. Thus the process of extinction is explained in terms of effect of need reduction (186).

mediate gradient of reinforcement. It is the failure to react to remote punishing consequences which is the basis for the persistence of behavior in the face of punishment. This analysis holds well for the behavior of Mowrer and Ullman's rats who, indeed, showed only a very limited capacity to learn in terms of temporally remote consequences. But what about non-integrative learning in humans capable of reactions to symbols? In face of this difficulty the authors unfortunately fall back on a question-begging argument: "The fact that certain habits or 'traits of character' may persist in the face of consistent punishment, raises a more difficult problem. One possibility is that punishment, which is observed, is offset by self-administered reward, which is not observed" (192, p. 85). But on this supposition the law of effect is incapable of disproof since whatever learning occurs is *ipso facto* considered the result of reward administered either visibly by others or invisibly and (often unobservably) by the self. Mowrer and Ullman probably sense this difficulty since they admit in the same paper that "there may be more to the problem than this and that the 'strength of the total ego' seems capable in ways which have not yet been clearly analyzed, of being mobilized in support of any single part (habit) for which the going is particularly hard" (p. 85).

In the symposium already referred to, Mowrer (185) has further clarified his view of the relation between the law of effect and "ego-processes." He does not believe that the type of learning which is described as "ego-involved" poses problems which the law of effect cannot answer. For the term "ego-involvement" he would substitute the term "interest"; for "interest" he would in turn substitute "emotional arousal"; and emotional arousal (e.g., fear) is a derived drive whose satisfaction is not in principle different from the reduction of the so-called primary, or biological drives. By such successive reductions Mowrer achieves the unification of all motives which may be effective in learning and staunchly defends the law of effect as the sole true law of learning. "Learning occurs when and only when a drive is reduced, a problem solved, a satisfaction derived, but . . . this satisfaction may stem from the reduction of either a primary or secondary drive" (185, p. 332).²⁴ It is the mediation of effect through symbolic processes,

²⁴ In his most recent statement Professor Mowrer (185a) has reversed his position to some extent. He now admits that the law of effect has failed as a universal monistic principle of learning and suggests that it may be necessary to assume two basic learning processes: (1) the processes whereby solutions to problems, i.e., ordinary habits, are acquired and (2) the processes whereby emotional learning or "conditioning" takes place. Habit formation or problem solving is mediated by the law of effect whereas conditioning or emotional learning is conceptualized as a process of "association" independent of

especially self-administered rewards, which distinguishes ego-involved learning from learning which is reinforced by the reduction of physiological needs.

An attempt rather similar to Mowrer's, though less orthodox in formulation, to bridge the gap between ego-psychology and the law of effect was made by Rice (221). Rice admits that the law cannot be held if it means that success or satisfaction leads to (a) a perpetuation of specific responses and/or (b) the repeated choice of the same specific goal object. When normal adults repeat one of these two elements, it is usually with a variation in the other, or both response and goal are varied. But even though either specific response sequence or goal or both may be varied, it may still be true that *something* about the activity is repeated, such as the "interest" that is involved. Rice suggests a reformulation of the law of effect. Success or satisfaction does not stamp in a specific stimulus-response connection but it does confirm the learner's *interest* in the general range of problems in which he has been successful. To the old question, *what* is it that success "stamps in," Rice then offers the answer that it is "interest." Interests themselves, even the processes to which the term ego refers, are acquired and perpetuated through the operation of this modified law of effect. Like Mowrer, Rice believes that symbolic processes (self-administered rewards) play an important role in the mediation of the effect. "That core of the act which constitutes the 'interest' is the feature of it which is most likely to be symbolized and repeatedly confirmed through approval of its symbol" (221 p. 316).

The papers of Mowrer and Rice make it clear that it is *logically* possible to reformulate the law of effect and to describe its operation in such a way as to make it universally applicable to animal and "ego-involved" learning alike. But as Allport's reply (4) to these papers shows, the psychologist who is primarily concerned with the analysis of ego-processes cannot be satisfied by such reductions and reformulations. Allport rejects the reduction of ego-processes to emotional arousal (secondary drives). Only certain emotional states are ego-involved and the two states should not be considered identical. Indeed, it may be the lack of emotional arousal which characterizes the smooth functioning of some ego-processes. Any law of learning based on "reduction of

pleasure or tension reduction. Mowrer also suggests that effect learning is mediated by the central nervous system whereas emotional learning is carried on by the autonomic nervous system. This reformulation considerably curtails the sphere of operation allotted to the law of effect. It remains to be seen whether this particular dichotomization will fare any better than previous attempts at mutually exclusive classifications of the learning process.

emotional tension" can not indiscriminately be applied to learning that proceeds from ego-interests. Allport staunchly maintains that the role of ego-processes is "irreducible." There are other basic phenomena of learning for which effect theory cannot account satisfactorily. The favorable influence of motor activity and active participation resists reduction to pleasure or satisfaction. Above all, the ubiquitous directive action of *interests* must be taken into account in any comprehensive theory of learning; "... learning proceeds because it is relevant to an interest system: it adds to knowledge, it differentiates items within the system, it broadens the range of equivalent stimuli Pleasure attending a single response, or even concatenations of response, is not decisive" (4, p. 346). The capacity to relate environmental events to one's interest system, to discriminate between relevant and irrelevant means, far transcends the operation of effect, however broadly conceived.

In the light of such considerations Allport considers the law of effect as a secondary principle in learning. He believes that satisfaction plays a decisive role only in those organisms and those situations in which ego-processes are not involved. In less mechanical forms of learning satisfaction loses importance and can be maintained only as a question-begging concept. The symbols which are invoked as mediators of self-reward are "vague molecular constructs that taper off into a kind of aimless triviality so far as explanatory power is concerned" (4, p. 344). At best, satisfaction or success serves as an *indicator* to the learner of how well he is adjusting to his problem. But though satisfaction helps the learner perceive the situation he uses this indication in a variable manner according to the pattern of interests that comprise his ego-structure.

CONCLUSIONS

In the course of this survey we have touched on learning behavior ranging from conditioned salivation to the development of interest systems in mature adults. To all these types of learning the law of effect has been applied by some, only to be rejected by others. In spite of an ever-growing volume of experimental and theoretical papers, agreement still seems to be a long way off. The golden jubilee of the law of effect will be celebrated in the not-too-distant future, but some of the basic issues for learning theory which the law of effect had hoped to answer still remain:

1. What is the agent responsible for reinforcement? Is it association by contiguity; an OK reaction; the reduction of tension, primary or derived; the satisfaction of an interest; the confirmation of a cognitive expectation; the approval of the ego?

2. What is it that is reinforced? Is it a neural bond; an S-R connection; a perceptual organization; an interest system; the ego?

3. What is the basic mechanism of reinforcement? Is it the lowering of synaptic resistances; cortical irradiation; or are the physiological processes which will find ultimate acceptance still unnamed?

Perhaps there is no one right answer among these alternatives. Probably more than one of these principles and additional ones that remain to be discovered will be needed in an integrated theory of learning. It is safe to say that at the present state of our knowledge the law of effect as a monistic principle explaining all learning has not been substantiated. As *one* of the behavioral facts of learning, it cannot be gainsaid.

BIBLIOGRAPHY

1. ADAMS, D. K. A restatement of the problem of learning. *Brit. J. Psychol.*, 1931, 22, 150-178.
2. ALLPORT, G. W. *Personality: a psychological interpretation*. New York: Holt, 1937.
3. ALLPORT, G. W. The ego in contemporary psychology. *Psychol. Rev.*, 1943, 50, 451-478.
4. ALLPORT, G. W. Effect: a secondary principle of learning. *Psychol. Rev.*, 1946, 53, 335-347.
5. ANDERSON, A. C. Time discrimination in the white rat. *J. comp. Psychol.*, 1932, 13, 27-55.
6. ANDERSON, A. C. Runway time and the goal gradient. *J. exp. Psychol.*, 1933, 16, 423-428.
7. ARPS, G. F. A preliminary report on work with knowledge versus work without knowledge. *Psychol. Rev.*, 1917, 24, 449-455.
8. BAIN, A. *The senses and the intellect*. (3d Ed.) London: Longmans, Green, 1868.
9. BALDWIN, J. M. *Mental development in the child and the race*. London: Macmillan, 1906.
10. BARLOW, M. C. The influence of electric shock in mirror tracing. *Amer. J. Psychol.*, 1933, 45, 478-487.
11. BERNARD, J. A note on non-informative shock. *J. exp. Psychol.*, 1941, 29, 407-412.
12. BERNARD, J. The specificity of the effect of shock on the acquisition and retention of motor and verbal habits. *J. exp. Psychol.*, 1942, 31, 69-78.
13. BERNARD, J., & GILBERT, R. W. The specificity of the effect of shock for error in maze learning with human subjects. *J. exp. Psychol.*, 1941, 28, 178-186.
14. BLODGETT, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, 4, 113-134.
15. BRANDT, H. The spread of the influence of reward to bonds remote in sequence and time. *Arch. Psychol.*, N. Y., 1935, 27, No. 180.
16. BROGDEN, W. J. The effect of frequency of reinforcement upon the level of conditioning. *J. exp. Psychol.*, 1939, 24, 419-431.
17. BROGDEN, W. J., LIPMAN, E. A., & CULLER, E. The role of incentive in conditioning and extinction. *Amer. J. Psychol.*, 1938, 51, 109-118.
18. BROWN, J. F., & FEDER, D. D. Thorndike's theory of learning as

- Gestalt psychology. *Psychol. Bull.*, 1934, 31, 426-437.
19. BROWN, W. On the law of effect. (Abstract.) *Psychol. Bull.*, 1935, 32, 528.
 20. BROWN, W. The positive effect of punishment. *J. comp. Psychol.*, 1939, 28, 17-22.
 21. BRUNSWIK, E. Probability as a determiner of rat behavior. *J. exp. Psychol.*, 1939, 25, 175-197.
 22. BUGELSKI, R. Extinction with and without sub-goal reinforcement. *J. comp. Psychol.*, 1938, 26, 121-134.
 23. BUGELSKI, R., & MILLER, N. E. A spatial gradient in the strength of avoidance responses. *J. exp. Psychol.*, 1938, 23, 494-505.
 24. BUNCH, M. E. The effect of electric shock as punishment for errors in human maze learning. *J. comp. Psychol.*, 1928, 8, 343-359.
 25. BUNCH, M. E. Certain effects of electric shock in learning a stylus maze. *J. comp. Psychol.*, 1935, 20, 211-242.
 26. BUNCH, M. E., & HAGMAN, E. P. The influence of electric shocks for errors in rational learning. *J. exp. Psychol.*, 1937, 21, 330-341.
 27. BUNCH, M. E., & MCTEER, F. P. The influence of punishment during learning upon retroactive inhibition. *J. exp. Psychol.* 1932, 15, 473-495.
 28. BUXTON, C. E. Latent learning and the goal gradient hypothesis. *Contrib. Psychol. Theory*, 1940, 2, No. 6.
 29. CARR, H. A. *Psychology: a study of mental activity*. New York: Longmans, Green, 1925.
 30. CARR, H. A. Teaching and learning. *J. genet. Psychol.*, 1930, 37, 189-218.
 31. CARR, H. A. The law of effect: a round table discussion: I. *Psychol. Rev.*, 1938, 45, 191-199.
 32. CARR, H. A., & FREEMAN, A. S. Time relationships in the formation of associations. *Psychol. Rev.*, 1919, 26, 465-473.
 33. CASON, H. The pleasure-pain theory of learning. *Psychol. Rev.*, 1932, 39, 440-466.
 34. CHATTOCK, A. P., & GRINDLEY, G. C. The effect of delayed reward on the maze performance of chickens. *Brit. J. Psychol.*, 1933, 23, 382-388.
 35. CLEMENTS, F. E. The effect of time on distance discrimination in the albino rat. *J. comp. Psychol.*, 1928, 8, 317-324.
 36. COLE, L. W. The relation of strength of stimulus to rate of learning. *J. Anim. Behav.*, 1911, 1, 111-124.
 37. COURTS, F. A. The alleged retroactive effect of visual stimuli subsequent to a given response. *J. exp. Psychol.*, 1937, 20, 144-154.
 38. COURTS, F. A. & WAGGONER, D. The effect of 'something happening' after a response. *J. exp. Psychol.*, 1938, 22, 383-387.
 39. COWLES, J. T. Food tokens as incentives for learning by chimpanzees. *Comp. Psychol. Monogr.*, 1937, 14, No. 71.
 40. COWLES, J. T., & NISSEN, H. W. Reward expectancy in delayed responses of chimpanzees. *J. comp. Psychol.*, 1937, 24, 345-358.
 41. CRAFTS, L. W., & GILBERT, R. W. The effect of punishment during learning upon retention. *J. exp. Psychol.*, 1934, 17, 73-84.
 42. CRESPI, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, 55, 467-517.
 - 42a. CRESPI, L. P. Amount of reinforcement and level of performance. *Psychol. Rev.*, 1944, 51, 341-357.
 43. CULLER, E. A. Recent advances in some concepts of conditioning. *Psychol. Rev.*, 1938, 45, 134-153.
 44. CULLER, E. A. The law of effect: a round table discussion: IV. *Psy-*

- chol. Rev.*, 1938, 45, 206-211.
45. DAND, A. 'Reward' and 'punishment' in learning. *Brit. J. Psychol.*, 1946, 36, 83-87.
 46. DASHIELL, J. F. *Fundamentals of objective psychology*. New York: Houghton Mifflin, 1928.
 47. DASHIELL, J. F. A survey and synthesis of learning theories. *Psychol. Bull.*, 1935, 32, 261-275.
 48. DASHIELL, J. F. The law of effect: a round table discussion: V. *Psychol. Rev.*, 1938, 45, 212-214.
 49. DAUB, C. T. The effect of doors on latent learning. *J. comp. Psychol.*, 1933, 15, 49-58.
 50. DECAMP, J. E. Relative distance as a factor in the white rat's selection of a path. *Psychobiol.*, 1920, 2, 245-253.
 51. DEISENROTH, C. E., & SPENCE, K. W. An investigation of the problem of latent learning in the white rat. (Abstract.) *Psychol. Bull.*, 1941, 38, 706.
 52. DENNY, M. R. The role of secondary reinforcement in a partial reinforcement learning situation. *J. exp. Psychol.*, 1946, 36, 373-389.
 53. DODSON, J. D. The relation of strength of stimulus to rapidity of habit formation in the kitten. *J. Anim. Behav.*, 1915, 5, 330-336.
 54. DODSON, J. D. Relative value of reward and punishment in habit formation. *Psychobiol.*, 1917, 1, 231-276.
 55. DODSON, J. D. The relative value of satisfying and annoying situations as motives in the learning process. *J. comp. Psychol.*, 1932, 14, 147-164.
 56. DORCUS, R. M., & GRAY, W. L. The effectiveness of food and electric shock in learning and retention by rats when applied at critical points in the maze. *J. comp. Psychol.*, 1932, 14, 191-218.
 57. DREW, G. C. The function of punishment in learning. *J. genet. Psychol.*, 1938, 52, 257-267.
 58. DREW, G. C. The speed of locomotion gradient in its relation to the goal gradient. *J. comp. Psychol.*, 1939, 27, 333-372.
 59. DUNLAP, K. A revision of the fundamental laws of habit formation. *Science*, 1928, 67, 360-362.
 60. EISENSON, J. Confirmation and information in reward and punishment. *Arch. Psychol.*, N. Y., 1935, 27, No. 181.
 61. ELLIOTT, M. H. The effect of change of reward on the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1928, 4, 19-30.
 62. ELLIOTT, M. H. The effect of appropriateness of reward on the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, 4, 91-98.
 63. ELLIOTT, M. H. The effect of change of "drive" on maze performance. *Univ. Calif. Publ. Psychol.*, 1929, 4, 185-188.
 64. ELLIOTT, M. H. Some determining factors in maze performance. *Amer. J. Psychol.*, 1930, 42, 315-317.
 65. ELLSON, D. G. The acquisition of a token-reward habit in dogs. *J. comp. Psychol.*, 1937, 24, 505-522.
 66. ELLWELL, J. L., & GRINDLEY, G. C. The effect of knowledge of results on learning and performance. I. A coordinated movement of the two hands. *Brit. J. Psychol.*, 1938, 29, 39-53.
 67. ESTES, W. K. Discrimination conditioning. I. A discriminative property of conditioned anticipation. *J. exp. Psychol.*, 1943, 32, 150-155.
 68. ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, 57, No. 3.
 69. ESTES, W. K., & SKINNER, B. F. Some quantitative properties of anxiety. *J. exp. Psychol.*, 1941, 29, 390-400.

70. EVERALL, E. E. Perseveration in the rat. *J. comp. Psychol.*, 1935, 19, 343-369.
71. FAIRLIE, C. W. The effect of shock at the "moment of choice" on the formation of a visual discrimination habit. *J. exp. Psychol.*, 1937, 21, 662-669.
72. FARBER, F. Spread of effect of reward and punishment in a multiple choice situation. *Proc. Iowa Acad. Sci.*, 1941, 48, 313-317.
73. FINAN, J. L. Quantitative studies in motivation. I. Strength of conditioning in rats under varying degrees of hunger. *J. comp. Psychol.*, 1940, 29, 119-134.
74. FINCH, G., & CULLER, E. Relation of forgetting to experimental extinction. *Amer. J. Psychol.*, 1935, 47, 656-662.
75. FINGER, F. W. Retention and subsequent extinction of a simple running response following varying conditions of reinforcement. *J. exp. Psychol.*, 1942, 31, 120-133.
76. FINGER, F. W. The effect of varying conditions of reinforcement upon a simple running response. *J. exp. Psychol.*, 1942, 30, 53-68.
77. FITTS, P. M. The effect of a large and a small reward as indicated by the resistance-to-extinction curve for the rat. (Abstract.) *Psychol. Bull.*, 1940, 37, 429-430.
78. FORLANO, G. *School learning with various methods of practice and reward*. Teach. Coll., Columbia Univ., Contr. Educ., No. 688. New York: Bureau of Publications, Teachers College, Columbia Univ. 1936.
79. GILBERT, R. W. The effect of non-informative shock upon maze learning and retention with human subjects. *J. exp. Psychol.*, 1936, 19, 456-466.
80. GILBERT, R. W. A further study of the effect of non-informative shock upon learning. *J. exp. Psychol.*, 1937, 20, 396-407.
81. GILBERT, R. W., & CRAFTS, L. W. The effect of signal for error upon maze learning and retention. *J. exp. Psychol.*, 1935, 18, 121-132.
82. GILLILAND, A. R. The effect of practice with and without knowledge of results in grading handwriting. *J. educ. Psychol.*, 1925, 16, 532-536.
83. GILLILAND, A. R. The law of effect in learning. *J. comp. Psychol.*, 1937, 24, 495-504.
84. GOODENOUGH, F. A note on Tolman's 'disproof' of Thorndike's law of effect. *J. exp. Psychol.*, 1933, 16, 459-462.
85. GRICE, G. R. An experimental study of the gradient of reinforcement in maze learning. *J. exp. Psychol.*, 1942, 30, 475-489.
86. GRINDLEY, G. C. Experiments on the influence of the amount of reward on learning in young chickens. *Brit. J. Psychol.*, 1929, 20, 173-180.
87. GURNEE, H. The effect of electric shock for right responses on maze learning. *J. exp. Psychol.*, 1938, 22, 354-364.
88. GUTHRIE, E. R. Conditioning as a principle of learning. *Psychol. Rev.*, 1930, 37, 412-428.
89. GUTHRIE, E. R. Reward and punishment. *Psychol. Rev.*, 1934, 41, 450-460.
90. GUTHRIE, E. R. *The psychology of learning*. New York: Harper, 1935.
91. GUTHRIE, E. R. *The psychology of human conflict*. New York: Harper, 1938.
92. GUTHRIE, E. R. The effect of outcome on learning. *Psychol. Rev.*, 1939, 46, 480-484.
93. GUTHRIE, E. R. Association and the law of effect. *Psychol. Rev.*, 1940, 47, 127-149.
94. GUTHRIE, E. R. Personality in terms of associative learning. In J. McV.

- Hunt (Ed.), *Personality and the behavior disorders*, Vol. I. New York: Ronald Press, 1944.
95. GUTHRIE, E. R., & HORTON, G. P. *Cats in a puzzle box*. New York: Rinehart, 1946.
 96. HAGGERTY, M. R. The laws of learning. *Psychol. Rev.*, 1913, 20, 411-422.
 97. HAIRE, M. A note concerning McCulloch's discussion of discrimination habits. *Psychol. Rev.*, 1939, 46, 298-303.
 98. HAIRE, M. Some factors influencing repetitive errors in discriminative learning. *J. comp. Psychol.*, 1939, 27, 79-91.
 99. HAIRE, M. Some experimental data relevant to field and associative theories of discriminative learning. *J. gen. Psychol.*, 1942, 26, 267-288.
 100. HAMILTON, E. L. The effect of delayed incentive on the hunger drive in the white rat. *Genet. Psychol. Monogr.*, 1929, 5, 131-207.
 101. HAMILTON, J. A., & KRESHEVSKY, I. Studies in the effect of shock on the behavior plasticity of the rat. *J. comp. Psychol.*, 1933, 16, 237-253.
 102. HANEY, G. W. The effect of familiarity on maze performance of albino rats. *Univ. Calif. Publ. Psychol.*, 1931, 4, 319-333.
 103. HERB, F. H. Latent learning—non-reward followed by food in blinds. *J. comp. Psychol.*, 1940, 29, 247-256.
 104. HERRICK, C. J. Awareness in the inductive solution of problems using words as stimuli. *Arch. Psychol.*, N. Y., 1939, 33, No. 234.
 105. HILGARD, E. R. The nature of the conditioned response: I. The case for and against stimulus substitution. *Psychol. Rev.*, 1936, 43, 366-385.
 106. HILGARD, E. R. The nature of the conditioned response: II. Alternatives to stimulus substitution. *Psychol. Rev.*, 1936, 43, 547-564.
 107. HILGARD, E. R. The relationship between the conditioned response and conventional learning experiments. *Psychol. Bull.*, 1937, 34, 61-102.
 108. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning*. New York: Appleton-Century, 1940.
 109. HILL, C. J. Goal gradient, anticipation and perseveration in compound trial-and-error learning. *J. exp. Psychol.*, 1939, 25, 566-585.
 110. HOBHOUSE, L. T. *Mind in evolution*. New York: Macmillan, 1901.
 111. HOGE, M. A., & STOCKING, R. J. A note on the relative value of punishment and reward as motives. *J. Anim. Behav.*, 1912, 2, 43-50.
 112. HOLLINGWORTH, H. L. *Psychology, its facts and principles*. New York: Appleton-Century, 1928.
 113. HOLLINGWORTH, H. L. Effect and affect in learning. *Psychol. Rev.*, 1931, 38, 153-159.
 114. HOLMES, S. J. *The evolution of animal intelligence*. New York: Holt, 1911.
 115. HONZIK, C. H., & TOLMAN, E. C. An experimental study of the effect of punishment on discrimination learning. (Abstract.) *Psychol. Bull.*, 1937, 34, 775.
 116. HONZIK, C. H., & TOLMAN, E. C. The action of punishment in accelerating learning. *J. comp. Psychol.*, 1938, 26, 187-200.
 117. HOVLAND, C. I. "Inhibition of reinforcement" and phenomena of experimental extinction. *Proc. Nat. Acad. Sci.*, 1936, 22, 430-433.
 118. HOVLAND, C. I. The generalization of conditioned response. IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses. *J. exp. Psychol.*, 1937, 21, 261-276.
 119. HULIN, W. S., & KATZ, D. A comparison of emphasis upon right and upon wrong responses in learn-

- ing. *J. exp. Psychol.*, 1935, 18, 638-642.
120. HULL, C. L. The goal gradient hypothesis and maze learning. *Psychol. Rev.*, 1932, 39, 25-43.
 121. HULL, C. L. Differential habituation to internal stimuli in the albino rat. *J. comp. Psychol.*, 1933, 16, 255-273.
 122. HULL, C. L. The rat's speed-of-locomotion gradient in the approach to food. *J. comp. Psychol.*, 1934, 17, 393-422.
 123. HULL, C. L. Special review of Thorndike, *The fundamentals of learning*. *Psychol. Bull.*, 1935, 32, 807-823.
 124. HULL, C. L. Mind, mechanism, and adaptive behavior. *Psychol. Rev.*, 1937, 44, 1-32.
 125. HULL, C. L. The goal gradient hypothesis applied to some 'field-force' problems in the behavior of young children. *Psychol. Rev.*, 1938, 45, 271-299.
 126. HULL, C. L. Outline of a systematic theory of learning. In *41st Yearbook Nat. Soc. Study Educ., Part II*. Bloomington: Public School Publishing Co., 1942.
 127. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
 128. HUMPHREYS, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, 25, 294-301.
 129. HUMPHREYS, L. G. The effect of random alternation and reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, 25, 141-158.
 130. HUMPHREYS, L. G. The strength of a Thorndikian response as a function of the number of practice trials. *J. comp. Psychol.*, 1943, 35, 101-110.
 131. IRWIN, R. W., KAUFMAN, K., PRIOR, G., & WEAVER, H. B. On 'learning without awareness of what is being learned'. *J. exp. Psychol.*, 1934, 17, 823-827.
 132. JAMES, W. *The principles of psychology*. New York: Holt, 1890.
 133. JENKINS, W. O. Studies in spread of effect. I. The bi-directional gradient in the performance of white rats in a linear maze. *J. comp. Psychol.*, 1943, 35, 41-56.
 134. JENKINS, W. O. Studies in spread of effect. II. The effect of increased motivation on the bi-directional gradient. *J. comp. Psychol.*, 1943, 35, 57-63.
 135. JENKINS, W. O. Studies in spread of effect. III. The effect of increased incentive on the bi-directional gradient. *J. comp. Psychol.*, 1943, 35, 65-72.
 136. JENKINS, W. O., & SHEFFIELD, F. D. Rehearsal and guessing habits as sources of the 'spread of effect.' *J. exp. Psychol.*, 1946, 36, 316-330.
 137. JENSEN, M. B. Punishment by electric shock as affecting performance in a raised finger maze. *J. exp. Psychol.*, 1934, 17, 65-72.
 138. JONES, F. N., & KERR, J. A comparison of the elimination of shock-right and shock-wrong blinds in a finger maze. *J. gen. Psychol.*, 1945, 33, 111-119.
 139. JONES, H. E. Trial and error learning with differential cues. *J. exp. Psychol.*, 1945, 35, 31-45.
 140. JUDD, C. H. Practice without knowledge of results. *Psychol. Monogr. Suppl.*, 1905, 7, 185-199.
 141. KAPPAUF, W. E., & SCHLOSBERG, H. Conditioned responses in the white rat. III. Conditioning as a function of the length of the period of delay. *J. genet. Psychol.*, 1937, 50, 27-45.
 142. KELLOGG, W. N. An eclectic view of some theories of learning. *Psychol. Rev.*, 1938, 45, 165-184.
 143. KELLOGG, W. N. Electric shock as a motivating stimulus in condition-

- ing experiments. *J. gen. Psychol.*, 1941, 25, 85-96.
144. KENDLER, H. H. Drive interaction: I. Learning as a function of the simultaneous presence of the hunger and thirst drives. *J. exp. Psychol.*, 1945, 35, 96-109.
 145. KENDLER, H. H. Drive interaction: II. Experimental analysis of the role of drive in learning theory. *J. exp. Psychol.*, 1945, 35, 188-198.
 146. KENDLER, H. H. The influence of simultaneous hunger and thirst drives upon the learning of two opposed spatial responses of the white rat. *J. exp. Psychol.*, 1946, 36, 212-220.
 147. KOCH, S., & DANIEL, W. J. The effect of satiation on the behavior mediated by a habit of maximum strength. *J. exp. Psychol.*, 1945, 35, 167-187.
 148. KOFFKA, K. *The growth of the mind*. New York: Harcourt, Brace, 1928.
 149. KRESHEVSKY, I. "Hypotheses" in rats. *Psychol. Rev.*, 1932, 39, 516-532.
 150. KRESHEVSKY, I. "Hypotheses" versus "chance" in the pre-solution period in sensory discrimination learning. *Univ. Calif. Publ. Psychol.*, 1932, 6, 27-44.
 151. KRESHEVSKY, I. The docile nature of hypotheses. *J. comp. Psychol.*, 1933, 15, 429-443.
 152. KRESHEVSKY, I. Hereditary nature of "hypotheses." *J. comp. Psychol.*, 1933, 16, 96-116.
 153. KRESHEVSKY, I. A note concerning the nature of discrimination learning in animals. *Psychol. Rev.*, 1937, 44, 97-104.
 154. KRESHEVSKY, I. A study of the continuity of the problem-solving process. *Psychol. Rev.*, 1938, 45, 107-133.
 155. KUO, Z. Y. The nature of unsuccessful acts and their order of elimination in animal learning. *J. comp. Psychol.*, 1922, 2, 1-27.
 156. LASHLEY, K. S. The mechanism of vision: I. A method for rapid analysis of pattern vision in the rat. *J. genet. Psychol.*, 1930, 37, 453-460.
 157. LASHLEY, K. S. An examination of the "continuity" theory as applied to discrimination learning. *J. gen. Psychol.*, 1942, 26, 241-265.
 158. LASHLEY, K. S., & WADE, MARJORIE. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, 53, 72-87.
 159. LEEPER, R. The role of motivation in learning; a study of the phenomenon of differential motivational control of the utilization of habits. *J. genet. Psychol.*, 1935, 46, 3-40.
 160. LEEPER, R. Dr. Hull's *Principles of behavior*. *J. genet. Psychol.*, 1944, 65, 3-52.
 161. LEWIN, K. Field theory and learning. In *41st Yearbook Nat. Soc. Study Educ.* Bloomington: Public School Publishing Co., 1942.
 162. LORGE, I. The effect of the initial chances for right responses upon the efficacy of intensified reward and intensified punishment. *J. exp. Psychol.*, 1933, 16, 362-373.
 163. LORGE, I. The efficacy of intensified reward and intensified punishment. *J. exp. Psychol.*, 1933, 16, 177-207.
 164. LORGE, I. Irrelevant rewards in animal learning. *J. comp. Psychol.*, 1936, 21, 105-128.
 165. LORGE, I., EISENSON, J., & EPSTEIN, B. Further experiments in the strength of connections where the connection is punished or rewarded or neither punished nor rewarded. *J. exp. Psychol.*, 1934, 17, 412-423.
 166. LORGE, I., & THORNDIKE, E. L. The comparative strengthening of a connection by one or more occurrences of it in cases where the con-

- nection was punished and was neither punished nor rewarded. *J. exp. Psychol.*, 1933, 16, 374-382.
167. LORGE, I., & THORNDIKE, E. L. The influence of delay in the after-effect of a connection. *J. exp. Psychol.*, 1935, 18, 186-194.
 168. McCULLOCH, T. L. Reply to a note on discrimination habits. *Psychol. Rev.*, 1939, 46, 304-307.
 169. McCULLOCH, T. L. The role of clasp- ing activity in adaptive behavior of of the infant chimpanzee. III. The mechanism of reinforcement. *J. Psychol.*, 1939, 7, 305-316.
 170. McCULLOCH, T. L., & PRATT, J. G. A study of the pre-solution period in weight discrimination by white rats. *J. comp. Psychol.*, 1934, 18, 271-290.
 171. MCGEOCH, J. A. *The psychology of human learning*. New York: Longmans, Green, 1945.
 172. McTEER, W. A study of certain fea- tures of punishment in serial learn- ing. *J. exp. Psychol.*, 1931, 14, 453-476.
 173. McTEER, W. Changes in grip tension following electric shock in mirror tracing. *J. exp. Psychol.*, 1933, 16, 735-742.
 174. MARX, M. H., & BUNCH, M. E. Error gradients around success in multi- ple choice learning. (Abstract.) *Psychol. Bull.*, 1942, 39, 593.
 175. MAY, M. A. *The social psychology of war and peace*. New Haven: Yale Univ. Press, 1943.
 176. MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the be- havior disorders*. New York: Ronald Press, 1944.
 177. MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univ. Press, 1941.
 178. MILLER, N. E., & MILES, W. R. Ef- fect of caffeine on the running speed of hungry, satiated, and frus- trated rats. *J. comp. Psychol.*, 1935, 20, 397-412.
 179. MITRANO, A. J. Principles of condi- tioning in human goal behavior. *Psychol. Monogr.*, 1939, 51, No. 230.
 180. MOWRER, O. H. Preparatory set (expectancy)—a determinant in motivation and learning. *Psychol. Rev.*, 1938, 45, 62-91.
 181. MOWRER, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, 46, 553-565.
 182. MOWRER, O. H. Anxiety-reduction and learning. *J. exp. Psychol.*, 1940, 27, 497-516.
 183. MOWRER, O. H. Preparatory set (expectancy)—some methods of measurement. *Psychol. Monogr.*, 1940, 52, No. 2.
 184. MOWRER, O. H. Motivation and learning in relation to the national emergency. *Psychol. Bull.*, 1941, 38, 421-431.
 185. MOWRER, O. H. The law of effect and ego psychology. *Psychol. Rev.*, 1946, 53, 321-334.
 - 185a. MOWRER, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem solv- ing." *Harvard Educ. Rev.*, 1947, 17, 102-148.
 186. MOWRER, O. H., & JONES, H. Ex- tinction and behavior variability as functions of effortfulness of task. *J. exp. Psychol.*, 1943, 33, 369-386.
 187. MOWRER, O. H., & JONES, H. Habit strength as a function of the pat- tern of reinforcement. *J. exp. Psychol.*, 1945, 35, 293-310.
 188. MOWRER, O. H., & KLUCKHOHN, C. Dynamic theory of personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald Press, 1944.
 189. MOWRER, O. H., & LAMOREAUX, R. R. Fear as an intervening vari- able in avoidance conditioning. *J. comp. Psychol.*, 1946, 39, 29-50.

190. MOWRER, O. H., & LAMOREAUX, R. R. Avoidance conditioning and signal duration—a study of secondary motivation and reward. *Psychol. Monogr.*, 1942, 54, No. 5.
191. MOWRER, O. H., RAYMAN, N. N., & BLISS, E. L. Preparatory set (expectancy)—an experimental demonstration of its central locus. *J. exp. Psychol.*, 1940, 26, 357–372.
192. MOWRER, O. H., & ULLMAN, A. D. Time as a determinant in integrative learning. *Psychol. Rev.*, 1945, 52, 61–90.
193. MUENZINGER, K. F. Motivation in learning. I. Electric shock for correct response in the visual discrimination habit. *J. comp. Psychol.*, 1934, 17, 267–277.
194. MUENZINGER, K. F. Motivation in learning. II. The function of electric shock for right and wrong responses in human subjects. *J. exp. Psychol.*, 1934, 17, 439–448.
195. MUENZINGER, K. F. The law of effect: a round table discussion. VI. *Psychol. Rev.*, 1938, 45, 215–218.
196. MUENZINGER, K. F., BERNSTONE, A. H., & RICHARDS, L. Motivation in learning. VIII. Equivalent amounts of electric shock for right and wrong responses in visual discrimination habit. *J. comp. Psychol.*, 1938, 26, 177–185.
197. MUENZINGER, K. F., & DOVE, C. C. Serial learning. I. Gradients of uniformity and variability produced by success and failure of single responses. *J. gen. Psychol.*, 1937, 16, 403–413.
198. MUENZINGER, K. F., DOVE, C. C., & BERNSTONE, A. H. Serial learning. II. The bi-directional goal gradient in the endless maze. *J. genet. Psychol.*, 1937, 50, 229–241.
199. MUENZINGER, K. F., & FLETCHER, F. M. Motivation in learning. VI. Escape from electric shock compared with hunger-food tension in the visual discrimination habit. *J. comp. Psychol.*, 1936, 22, 79–91.
200. MUENZINGER, K. F., & FLETCHER, F. M. Motivation in learning. VII. The effect of an enforced delay at the point of choice in the visual discrimination habit. *J. comp. Psychol.*, 1937, 23, 383–392.
201. MUENZINGER, K. F., & NEWCOMB, H. Motivation in learning. III. A bell signal compared with electric shock for right and wrong responses in the visual discrimination habit. *J. comp. Psychol.*, 1935, 20, 85–93.
202. MUENZINGER, K. F., & NEWCOMB, H. Motivation in learning. V. The relative effectiveness of jumping a gap and crossing an electric grid in a visual discrimination habit. *J. comp. Psychol.*, 1936, 21, 95–104.
203. MUENZINGER, K. F., & VINE, D. O. Motivation in learning. IX. The effect of interposed obstacles in human learning. *J. exp. Psychol.*, 1941, 29, 67–74.
204. MUENZINGER, K. F., & WOOD, A. Motivation in learning. IV. The function of punishment as determined by its temporal relation to the act of choice in the visual discrimination habit. *J. comp. Psychol.*, 1935, 20, 95–106.
205. MUNN, N. L. An apparatus for testing visual discrimination in animals. *J. genet. Psychol.*, 1931, 39, 342–358.
206. NI, CHUNG-FANG. An experimental study of the influence of punishment for errors during learning upon retention. *J. comp. Psychol.*, 1934, 17, 279–301.
207. NISSEN, H. W., & CRAWFORD, M. P. A preliminary study of food-sharing behavior in young chimpanzees. *J. comp. Psychol.*, 1936, 22, 383–419.
208. NISSEN, H. W., & ELDER, J. H. The influence of the amount of incentive on delayed response performance.

- ance of chimpanzees. *J. genet. Psychol.*, 1935, 47, 49-72.
209. PERIN, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, 30, 93-113.
 210. PERIN, C. T. A quantitative investigation of the delay-of-reinforcement gradient. *J. exp. Psychol.*, 1943, 32, 37-51.
 211. PERIN, C. T. The effect of delayed reinforcement upon the differentiation of bar responses in white rats. *J. exp. Psychol.*, 1943, 32, 95-109.
 212. PETERSON, J. Forced adjustment versus association in constructive learning and thinking. *Amer. J. Psychol.*, 1927, 39, 264-282.
 213. PETERSON, J. A review of *The fundamentals of learning* by E. L. Thorndike. *Amer. J. Psychol.*, 1935, 47, 707-710.
 214. PORTER, E. H., JR., & HALL, C. S. A further investigation of the role of emphasis in learning. *J. exp. Psychol.*, 1938, 22, 377-382.
 215. PYLE, W. H. Origin of the pleasure-pain theory of learning. *J. educ. Psychol.*, 1933, 24, 303-304.
 216. RAZRAN, G. H. S. The law of effect or the law of qualitative conditioning. *Psychol. Rev.*, 1939, 46, 445-463.
 217. REED, H. B. An experiment on the law of effect in learning the maze by human subjects. *J. educ. Psychol.*, 1935, 26, 695-700.
 218. REXROAD, C. N. Administering electric shock for inaccuracy in continuous multiple-choice reactions. *J. exper. Psychol.*, 1926, 9, 1-18.
 219. REXROAD, C. N. Outline of the conditions under which learning occurs. *Psychol. Rev.*, 1932, 39, 174-183.
 220. REYNOLDS, B. A repetition of the Blodgett experiment on latent learning. *J. exp. Psychol.*, 1945, 35, 504-516.
 221. RICE, P. B. The ego and the law of effect. *Psychol. Rev.*, 1946, 53, 307-320.
 222. ROBERTS, W. H. The effect of delayed feeding on white rats in a problem cage. *J. genet. Psychol.*, 1930, 37, 35-58.
 223. ROCK, R. T., JR. *The influence upon learning of the quantitative variation of after-effect*. Teach. Coll., Columbia Univ., Contr. Educ., No. 650. New York: Bureau of Publications, Teachers College, Columbia Univ., 1935.
 224. ROCK, R. T., JR. Knowledge of results of responses compared with after-effects following each response in a learning situation. (Abstract.) *Psychol. Bull.*, 1936, 33, 613.
 225. ROSS, C. C. An experiment in motivation. *J. educ. Psychol.*, 1927, 18, 337-346.
 226. SAMS, C. F., & TOLMAN, E. C. Time discrimination in white rats. *J. comp. Psychol.*, 1925, 5, 255-263.
 227. SANDIFORD, P. Connectionism. In *41st Yearbook Nat. Soc. Study Educ., Part II*. Bloomington: Public School Publ. Co., 1942.
 228. SCHLOSBERG, H. The relationship between success and the laws of conditioning. *Psychol. Rev.*, 1937, 44, 379-394.
 229. SEASHORE, H., & BAVELAS, A. The functioning of knowledge of results in Thorndike's line-drawing experiment. *Psychol. Rev.*, 1941, 48, 155-164.
 230. SEWARD, J. P. An experimental study of Guthrie's theory of reinforcement. *J. exp. Psychol.*, 1942, 30, 247-256.
 231. SEWARD, J. P. Reinforcement in terms of association. *Psychol. Rev.*, 1943, 50, 187-202.
 232. SEWARD, J. P. A theoretical derivation of latent learning. *Psychol. Rev.*, 1947, 54, 83-98.
 233. SIEGEL, P. S. The role of absolute

- initial response strength in simple trial-and-error learning. *J. exp. Psychol.*, 1945, 35, 199-205.
234. SILLECK, S. B., JR., & LAPHA, C. W. The relative effectiveness of emphasis upon right and wrong responses in human maze learning. *J. exp. Psychol.*, 1937, 20, 195-201.
 235. SIMLEY, O. A. The relation of subliminal to supraliminal learning. *Arch. Psychol.*, N. Y., 1933, 22, No. 146.
 236. SKINNER, B. F. Two types of conditioned reflex and a pseudo type. *J. gen. Psychol.*, 1935, 12, 66-77.
 237. SKINNER, B. F. *The behavior of organisms: an experimental analysis*. New York: Appleton-Century, 1938.
 238. SPENCE, K. W. The order of eliminating blinds in maze learning by the rat. *J. comp. Psychol.*, 1932, 14, 9-27.
 239. SPENCE, K. W. Visual acuity and its relation to brightness in chimpanzee and man. *J. comp. Psychol.*, 1934, 18, 333-361.
 240. SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, 43, 427-449.
 241. SPENCE, K. W. Analysis of the formation of visual discrimination habits of chimpanzees. *J. comp. Psychol.*, 1937, 23, 77-100.
 242. SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, 44, 430-444.
 243. SPENCE, K. W. Gradual versus sudden solution of discrimination problems by chimpanzees. *J. comp. Psychol.*, 1938, 25, 213-224.
 244. SPENCE, K. W. Continuous versus non-continuous interpretations of discrimination learning. *Psychol. Rev.*, 1940, 47, 271-288.
 245. SPENCE, K. W. An experimental test of the continuity and non-continuity theories of discrimination learning. *J. exp. Psychol.*, 1945, 35, 253-266.
 246. SPENCE, K. W. Theoretical interpretations of learning. In F. A. Moss (Ed.), *Comparative psychology*. New York: Prentice-Hall, 1946.
 247. SPENCE, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, 54, 1-8.
 248. SPENCE, K. W., & LIPPITT, R. O. "Latent" learning of a simple maze problem with relevant needs satiated. (Abstract.) *Psychol. Bull.*, 1940, 37, 429.
 249. SPENCE, K. W., & LIPPITT, R. O. An experimental test of the sign-gestalt theory of trial-and-error learning. *J. exp. Psychol.*, 1946, 36, 491-502.
 250. SPENCER, H. *Principles of psychology*. London: Williams & Norgate, 1872.
 251. STEPHENS, J. M. A mechanical explanation of the law of effect. *Amer. J. Psychol.*, 1929, 41, 422-431.
 252. STEPHENS, J. M. *The influence of different stimuli upon preceding bonds*. Teach. Coll., Columbia Univ., Contr. Educ., No. 493. New York: Bureau of Publications, Teachers College, Columbia Univ., 1931.
 253. STEPHENS, J. M. A change in the interpretation of the law of effect. *Brit. J. Psychol.*, 1934, 24, 266-275.
 254. STEPHENS, J. M. Further notes on punishment and reward. *J. genet. Psychol.*, 1934, 44, 464-472.
 255. STEPHENS, J. M. The influence of punishment on learning. *J. exp. Psychol.*, 1934, 17, 536-555.
 256. STEPHENS, J. M. The influence of symbolic punishment and reward upon strong and weak associations. *J. gen. Psychol.*, 1941, 25, 177-185.
 257. STEPHENS, J. M. Expectancy vs. effect—substitution as a general principle of reinforcement. *Psychol. Rev.*, 1942, 49, 102-116.
 258. STEPHENS, J. M. Is habit-formation always mediated by cognition-

- formation? *Psychol. Rev.*, 1943, 50, 430-437.
259. STEPHENS, J. M., & BAER, J. A. The influence of punishment on learning when the opportunity for inner repetition is reduced. *J. genet. Psychol.*, 1937, 51, 209-217.
 260. STONE, G. R. The effect of negative incentive on serial learning. I. The spread of variability under electric shock. *J. exp. Psychol.*, 1946, 36, 137-142.
 261. SZYMANSKI, J. S. Modification of the innate behavior of cockroaches. *J. Anim. Behav.*, 1912, 2, 81-90.
 262. THOMPSON, M. E., & DOVE, C. C. A further study of the bi-directional goal-gradient in the endless maze. *J. exp. Psychol.*, 1943, 33, 447-456.
 263. THORNDIKE, E. L. Animal intelligence: an experimental study of the associative process in animals. *Psychol. Monogr.*, 1898, 2, No. 8.
 264. THORNDIKE, E. L. *Animal intelligence: experimental studies*. New York: Macmillan, 1911.
 265. THORNDIKE, E. L. *Educational psychology, Vol. I. The original nature of man*. New York: Teachers College, Columbia Univ., 1913.
 266. THORNDIKE, E. L. The law of effect. *Amer. J. Psychol.*, 1927, 39, 212-222.
 267. THORNDIKE, E. L. *Human learning*. New York: Century, 1931.
 268. THORNDIKE, E. L. Rewards and punishment in animal learning. *Comp. Psychol. Monogr.*, 1932, 8, No. 65.
 269. THORNDIKE, E. L., et al. *The fundamentals of learning*. New York: Teachers College, Columbia Univ., 1932.
 270. THORNDIKE, E. L. *An experimental study of rewards*. Teach. Coll., Columbia Univ., Contr. Educ., No. 580. New York: Bureau of Publications, Teachers College, Columbia Univ., 1933.
 271. THORNDIKE, E. L. The influence of irrelevant rewards. *J. educ. Psychol.*, 1933, 24, 1-15.
 272. THORNDIKE, E. L. A proof of the law of effect. *Science*, 1933, 77, 173-175.
 273. THORNDIKE, E. L. A theory of the action of the after-effects of a connection upon it. *Psychol. Rev.*, 1933, 40, 434-439.
 274. THORNDIKE, E. L. The spread of the influence of reward to connections irrelevant to the learner's purposes. *J. genet. Psychol.*, 1934, 44, 428-435.
 275. THORNDIKE, E. L. The direct action of rewards upon mental connections and their indirect action via the stimulation of inner equivalents of the connections. *J. exp. Psychol.*, 1935, 18, 91-96.
 276. THORNDIKE, E. L. *The psychology of wants, interests, and attitudes*. New York: Appleton-Century, 1935.
 277. THORNDIKE, E. L. The law of effect: a round table discussion: III. *Psychol. Rev.*, 1938, 45, 204-205.
 278. THORNDIKE, E. L. *Human nature and the social order*. New York: Macmillan, 1940.
 279. THORNDIKE, E. L. *Man and his works*. Cambridge: Harvard Univ. Press, 1943.
 280. THORNDIKE, E. L., & FORLANO, G. The influence of increase and decrease of the amount of reward upon the rate of learning. *J. educ. Psychol.*, 1933, 24, 401-411.
 281. THORNDIKE, E. L., & LORGE, I. The influence of relevance and belonging. *J. exp. Psychol.*, 1935, 18, 574-584.
 282. THORNDIKE, E. L., & ROCK, R. T., JR. Learning without awareness of what is being learned or intent to learn it. *J. exp. Psychol.*, 1934, 17, 1-19.
 283. TILTON, J. W. The effect of "right" and "wrong" upon the learning of nonsense syllables in multiple choice arrangement. *J. educ. Psychol.*, 1939, 30, 95-115.

284. TILTON, J. W. The relative importance of success and failure in learning as related to certain individual differences. *J. educ. Psychol.*, 1943, **34**, 176-180.
285. TILTON, J. W. "Effect" as determined by initial strength of response. *J. gen. Psychol.*, 1944, **31**, 277-281.
286. TILTON, J. W. Gradients of effect. *J. genet. Psychol.*, 1945, **66**, 3-19.
287. TOLMAN, E. C. *Purposive behavior in animals and men*. New York: Century, 1932.
288. TOLMAN, E. C. Sign-gestalt or conditioned reflex? *Psychol. Rev.*, 1933, **40**, 246-255.
289. TOLMAN, E. C. Theories of learning. In F. A. Moss (Ed.), *Comparative psychology*. New York: Prentice-Hall, 1934.
290. TOLMAN, E. C. The acquisition of string-pulling by rats—conditioned response or sign-gestalt? *Psychol. Rev.*, 1937, **44**, 195-211.
291. TOLMAN, E. C. The determiners of behavior at a choice point. *Psychol. Rev.*, 1938, **45**, 1-41.
292. TOLMAN, E. C. The law of effect: a round table discussion: II. *Psychol. Rev.*, 1938, **45**, 200-203.
293. TOLMAN, E. C., & BRUNSWIK, E. The organism and the causal texture of the environment. *Psychol. Rev.*, 1935, **42**, 43-77.
294. TOLMAN, E. C., HALL, C. S., & BRETNALL, E. P. A disproof of the law of effect and a substitution of the laws of emphasis, motivation, and disruption. *J. exp. Psychol.*, 1932, **15**, 601-614.
295. TOLMAN, E. C., & HONZIK, C. H. Introduction and removal of reward and maze performance in rats. *Univ. Calif. Publ. Psychol.*, 1930, **4**, 257-275.
296. TOMKINS, S. S. An analysis of the use of electric shock with human subjects. *J. Psychol.*, 1943, **15**, 285-297.
297. TROLAND, L. T. *The fundamentals of human motivation*. New York: Van Nostrand, 1928.
298. TROWBRIDGE, M. H., & CASON, H. An experimental study of Thorndike's theory of learning. *J. gen. Psychol.*, 1932, **7**, 245-258.
299. TUCKMAN, J. *The influence of varying amounts of punishment on mental connections*. Teach. Coll., Columbia Univ., Contr. Educ., No. 590. New York: Bureau of Publications, Teachers College, Columbia Univ., 1933.
300. TURNER, C. H. An experimental investigation of an apparent reversal of light responses of the roach. *Biol. Bull.*, 1912, **23**, 371-386.
301. VALENTINE, R. The effects of punishment for errors on the maze learning of rats. *J. comp. Psychol.*, 1930, **10**, 35-53.
302. VAUGHN, J., & DISERENS, C. M. The relative effects of various intensities of punishment on learning and efficiency. *J. comp. Psychol.*, 1930, **10**, 55-66.
303. WAITS, J. V. The law of effect in the retained situation. *Arch. Psychol.*, N. Y., 1937, **30**, No. 208.
304. WALLACE, S. R., JR., BLACKWELL, M. G., JR., & JENKINS, I. Pre-reward and post-reward performance in the "latent learning" of an elevated maze. (Abstract.) *Psychol. Bull.*, 1941, **38**, 694.
305. WALLACH, H., & HENLE, MARY. An experimental analysis of the law of effect. *J. exp. Psychol.*, 1941, **28**, 340-349.
306. WALLACH, H., & HENLE, MARY. A further study of the function of reward. *J. exp. Psychol.*, 1942, **30**, 147-160.
307. WARDEN, C. J., & AYLESWORTH, M. The relative value of rewards and punishment in the formation of a visual discrimination habit in the white rat. *J. comp. Psychol.*, 1927, **7**, 117-127.

308. WARDEN, C. J., & DIAMOND, S. A preliminary study of the effect of delayed punishment in learning in the white rat. *J. genet. Psychol.*, 1931, 39, 455-462.
309. WARDEN, C. J., & HAAS, E. L. The effect of short intervals of delay in feeding upon speed of maze learning. *J. comp. Psychol.*, 1927, 7, 107-116.
310. WARDEN, C. J. & LUBOW, L. Effect of performance without reward on the retention of the maze habit in the white rat. *J. genet. Psychol.*, 1942, 60, 321-328.
311. WARNER, L. H. The association span of the white rat. *J. genet. Psychol.*, 1932, 41, 57-90.
312. WATERS, R. H. The law of effect as a principle of learning. *Psychol. Bull.*, 1934, 31, 408-425.
313. WATSON, J. B. The effect of delayed feeding upon learning. *Psychobiol.*, 1917, 1, 51-60.
314. WHITE, R. K. The completion hypothesis and reinforcement. *Psychol. Rev.*, 1936, 43, 396-404.
315. WHITE, R. K. The case for the Tolman-Lewin interpretation of learning. *Psychol. Rev.*, 1943, 50, 157-186.
316. WHITING, J. W. M. *Becoming a Kwoma*. New Haven: Yale Univ. Press, 1941.
317. WILLIAMS, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, 23, 506-522.
318. WOLFE, J. B. The effect of delayed reward upon learning in the white rat. *J. comp. Psychol.*, 1934, 17, 1-22.
319. WOLFE, J. B. Effectiveness of token-rewards for chimpanzees. *Comp. Psychol. Monogr.*, 1936, 12, No. 60.
320. WOLFE, J. B., & KAPLON, M. Effect of amount of reward and consummative activity on learning in chickens. *J. comp. Psychol.*, 1941, 31, 353-361.
321. WOOD, A. B. A comparison of delayed reward and delayed punishment in the formation of a brightness discrimination habit in the chick. *Arch. Psychol.*, N. Y., 1933, 24, No. 157.
322. WYLIE, H. H. An experimental study of transfer of response in the white rat. *Behav. Monogr.*, 1919, 3, No. 66.
323. YARBOROUGH, J. U. The influence of the time interval upon the rate of learning in the white rat. *Psychol. Monogr.*, 1921, 30, No. 135.
324. YERKES, R. M. *The dancing mouse*. New York: Macmillan, 1907.
325. YERKES, R. M. The intelligence of earthworms. *J. Anim. Behav.*, 1912, 2, 332-352.
326. YERKES, R. M., & DODSON, J. D. The relation of strength of stimulus to rapidity of habit formation. *J. comp. Neurol. and Psychol.*, 1908, 18, 459-482.
327. YOSHIOKA, J. G. Weber's law in the discrimination of maze distance by the white rat. *Univ. Calif. Publ. Psychol.*, 1929, 4, 155-184.
328. YOUNG, P. T. *Motivation of behavior*. New York: Wiley, 1936.
329. YOUTZ, R. E. P. Reinforcement, extinction and spontaneous recovery in a non-Pavlovian reaction. *J. exp. Psychol.*, 1938, 22, 305-318.
330. ZENER, K. E., & MCCURDY, H. G. Analysis of motivational factors in hunger upon conditioned, unconditioned, and spontaneous salivary secretion. *J. Psychol.*, 1939, 8, 321-350.
331. ZIRKLE, G. A. Success and failure in serial learning. I. The Thorndike effect. *J. exp. Psychol.*, 1946, 36, 230-236.
332. ZIRKLE, G. A. Success and failure in serial learning. II. Isolation and the Thorndike effect. *J. exp. Psychol.*, 1946, 36, 302-315.

DISCUSSION OF SNYDER'S "THE PRESENT STATUS OF PSYCHOTHERAPEUTIC COUNSELING"*

J. L. MORENO

Beacon, N. Y.

Psychiatrists have frequently claimed that psychologists are not "fit" to practice psychotherapy, but their attitude is changing in recent years. Psychologists, in turn, often assert that they are superior to psychiatrists in the use of scientific methods. However, an objective approach and understanding—from both sides of the fence—is of the utmost importance if we are ever to attain a united front among the psychotherapists. By articles like the one of Snyder the existing chaos in psychotherapeutic counseling will be increased and the cleavage between psychiatrists and psychologists further widened.** The author frequently stresses that psychiatrists are commonly inclined to be unscientific, subjective and intuitive in approach, whereas psychologists believe in experimental methods and scientific validation; that psychiatrists indulge in esoteric terminology, but that psychologists use terms which are universal, that is, accepted everywhere by the scientific fraternity. Such bias, which runs intermittently throughout the review comes to a head in the statements directed against Freud and Moreno. In the chapter on psychoanalysis he says:

One difficulty for the nonpsychoanalytically trained reader is that the vocabulary of psychoanalysis is so esoteric that it frequently has little meaning. . . . the majority of psychiatrists in this country and almost every psychologist are more likely to agree with Blanchette (?) that psychoanalysis is a medical psychology without adequate roots in or connections with scientific medicine or scientific psychology.

In the chapter on psychodrama he says:

Moreno himself is unscientific and intuitive in approach. . . . Moreno postulates relationships and behavior patterns which are explainable only in terms of his esoteric system. In this regard his method is similar to Freudian psychoanalysis (p. 333).

It is common knowledge that psychoanalysis is rooted in French psychiatry, the school of Charot. If it added some original ideas to medical psychology, this in itself should not make it unscientific. To

* SNYDER, W. U. The present status of psychotherapeutic counseling. *Psychol. Bull.*, 1947, 44, 297-386.

** Further elaborated in: Moreno, J. L. The chaos in therapeutic counselling, *Social try*, 1947, 1, No. 3.

treat Freud in an article on therapeutic counseling as if he would be a dead dog is ridiculous. He was a great empiricist, he formulated his hypotheses with care and discarded them when new observations seemed to warrant it. Because he was pre-experimental, this does not mean that he was unscientific; as the psychology of the individual was in its infancy when he approached it, he was as scientific as its stage of development permitted. Unfortunately, the words "scientific" and "esoteric" are used by Snyder in a moralizing, almost schoolmasterly way. All scientific terms have once been esoteric and many are now universally accepted. Should one hesitate to introduce new ideas and coin new terms because they are bound to be considered esoteric by our contemporaries? It should not be permitted that fear of new ideas and resentment of originality raise their head under the guise of a narrow concept of science and of how new ideas emerge.

The treatment which psychodrama and sociometry receive at the hand of Snyder is intuitive and esoteric, to use his own terminology. If he would have studied the subject more thoroughly he would have learned that the outstanding development in counseling in recent years is closely linked to sociometric and psychodramatic methods. It has nothing to do with the controversy "directive-nondirective" which psychiatrists have been carrying on for more than fifty years. In "sociometric counseling" the counselor may be anonymous; the counselor is not only nondirective but may be altogether *absent* from the scene when the therapeutic choice process takes place. The sociogram, mapping the emotional currents operating among the members of a group, becomes the objective reference upon which his recommendations as counselor are based. The process of involvement is in principle removed "from the counselor-client relationship to the client-client relationships.

Psychodramatic counseling is more complex; practically all problems which appear singly in each of the current psychotherapies reappear in a psychodramatic session magnified or in new combinations. Because there are many types of problems which the clients bring to us there are many versions of psychodrama. There has not been found a single method which can be applied universally. The oldest and most popular form of psychodrama is the *self directed* form, in which the self of the client directs the production exclusively, director and auxiliary egos assisting him in the production if he wishes. In another version of psychodrama the director is outside of the situation, the therapeutic process takes place between the two or more clients involved in the conflict.

There can be no greater contrast than the one between the non-directive interview and self-directed psychodrama production. Non-directive puts a premium on a *minimum of self expression of the counselor*; psychodrama puts a premium on the *maximum of self expression of the client*. The word nondirective is often misleading. It is a negative term, implying that the spontaneity of the client should be permitted to express itself without interference from any source. Non-directive writers talk often the language of spontaneity in reverse; they are worried about making the counselor as unstimulating to the client as possible but not worried about whether the subject is stimulated to be himself. Various interviewers have been observed in practice; it was found that, however cautious they may be in their questionings, on the action level they assumed various attitudes which appealed to the subject in uncontrollable ways. One counselor may have a fatherly way of taking the role, smiling approvingly; the other a businesslike way; the third may be withdrawn and indifferent; all three may use the same technique of questioning but they have a different effect upon the client. It was evident that the very physical appearance of the counselor was bound to influence the client; it has, as we sociometrists phrase it, a tele effect.

Mr. Snyder refers to one of my papers "Scientific foundations of group psychotherapy." He is unaware that this article is just a summary, that it refers back to fifteen years of research in which several hundred workers have collaborated, contained in nine volumes of *Sociometry*, books like Moreno's *Who Shall Survive?*, Jennings' *Leadership and Isolation*, monographs like "Sociometric statistics of social configurations," "The measurement of sociometric status, structure and development," etc. The climax of ignorance for the meaning of psychodrama and group psychotherapy comes forth in Snyder's statement: "In evaluating psychodrama it is perhaps of significance to note that Moreno himself performs for the public at two weekly sessions in New York where admission is charged" (p. 334).^{*} This is obviously a derogatory remark. Snyder seems to believe that it is unworthy of a psychotherapist, perhaps also unscientific as well as undignified, to give sessions "for the public."^{*} The fact is that the most revolutionary change in therapeutic counseling since psychoanalysis is linked with *open* sessions for the public. My giving of open sessions

^{*} At the New York Psychodramatic Institute a ten weeks' seminar, one session weekly, is given by me as is customary for experts in a specialty. The fee for the entire seminar is \$15.00. The fee for a single admission is \$1.65. The public consists of students of the universities and colleges in and around New York.

had enormous consequences because it broke with two sacred clichés, first that of the counselor being closeted in a secret chamber with the client and no other client or counselor sharing in the experiences evoked in the session; it was and had to remain strictly confidential. The public session opened the door wide for the participation of larger and smaller groups in a common therapeutic experience. Second, not making any record during the session which was to remain sacred and discreet for the sake of the client. But once the privacy of the one-to-one therapeutic relationship was broken it became a foregone conclusion that simultaneous observation and recording of the therapeutic process should be permissible. This marked the way towards the objectification of the interview. The clinical advance inaugurated by the public session meant automatically also an advance in objectively recording and analyzing its operations. The principle of electric recording taking place simultaneously with the therapeutic process was announced by me seventeen years ago and followed since then throughout my writings and frequently practiced—it is curious therefore that Snyder should give the nondirective counselor priority in this:

Reactions witnessed by the psychologist and revelations given by any individual during a course of interviews, casual or planned, are, at least from the point of view of cooperative, controllable research, of little value since they are after the event merely memory impressions of the observer. The multiform interpretations offered by the subjectivists in psychology are without proper demonstration and reconsideration as long as they do not conserve the moment. It was suggested frequently therefore that "a talking machine should photograph the process" and that we should make systematic use of this machinery of personality recording.*

Psychodramatic procedure, being an offshoot of sociometry, has to be viewed and evaluated with sociometric methodology and experimentation as a background. Terms like *sociodynamic effect*, *social atom* and *tele* can appear esoteric and intuitive only to one who has not acquainted himself with the methods by which they have been explored.

* MORENO, J. L. *Application of the group method to classification*. New York: The National Committee on Prisons and Prison Labor, 1931, p. 16. (Current edition: *Group method and group psychotherapy*. New York: Beacon House, 1946, p. 16.)

There can be no greater contrast than the one between the non-directive interview and self-directed psychodrama production. Non-directive puts a premium on a *minimum of self expression of the counselor*; psychodrama puts a premium on the *maximum of self expression of the client*. The word nondirective is often misleading. It is a negative term, implying that the spontaneity of the client should be permitted to express itself without interference from any source. Non-directive writers talk often the language of spontaneity in reverse; they are worried about making the counselor as unstimulating to the client as possible but not worried about whether the subject is stimulated to be himself. Various interviewers have been observed in practice; it was found that, however cautious they may be in their questionings, on the action level they assumed various attitudes which appealed to the subject in uncontrollable ways. One counselor may have a fatherly way of taking the role, smiling approvingly; the other a businesslike way; the third may be withdrawn and indifferent; all three may use the same technique of questioning but they have a different effect upon the client. It was evident that the very physical appearance of the counselor was bound to influence the client; it has, as we sociometrists phrase it, a tele effect.

Mr. Snyder refers to one of my papers "Scientific foundations of group psychotherapy." He is unaware that this article is just a summary, that it refers back to fifteen years of research in which several hundred workers have collaborated, contained in nine volumes of *Sociometry*, books like Moreno's *Who Shall Survive?*, Jennings' *Leadership and Isolation*, monographs like "Sociometric statistics of social configurations," "The measurement of sociometric status, structure and development," etc. The climax of ignorance for the meaning of psychodrama and group psychotherapy comes forth in Snyder's statement: "In evaluating psychodrama it is perhaps of significance to note that Moreno himself performs for the public at two weekly sessions in New York where admission is charged" (p. 334).^{*} This is obviously a derogatory remark. Snyder seems to believe that it is unworthy of a psychotherapist, perhaps also unscientific as well as undignified, to give sessions "for the public."^{*} The fact is that the most revolutionary change in therapeutic counseling since psychoanalysis is linked with *open* sessions for the public. My giving of open sessions

^{*} At the New York Psychodramatic Institute a ten weeks' seminar, one session weekly, is given by me as is customary for experts in a specialty. The fee for the entire seminar is \$15.00. The fee for a single admission is \$1.65. The public consists of students of the universities and colleges in and around New York.

had enormous consequences because it broke with two sacred clichés, first that of the counselor being closeted in a secret chamber with the client and no other client or counselor sharing in the experiences evoked in the session; it was and had to remain strictly confidential. The public session opened the door wide for the participation of larger and smaller groups in a common therapeutic experience. Second, not making any record during the session which was to remain sacred and discreet for the sake of the client. But once the privacy of the one-to-one therapeutic relationship was broken it became a foregone conclusion that simultaneous observation and recording of the therapeutic process should be permissible. This marked the way towards the objectification of the interview. The clinical advance inaugurated by the public session meant automatically also an advance in objectively recording and analyzing its operations. The principle of electric recording taking place simultaneously with the therapeutic process was announced by me seventeen years ago and followed since then throughout my writings and frequently practiced—it is curious therefore that Snyder should give the nondirective counselor priority in this:

Reactions witnessed by the psychologist and revelations given by any individual during a course of interviews, casual or planned, are, at least from the point of view of cooperative, controllable research, of little value since they are after the event merely memory impressions of the observer. The multiform interpretations offered by the subjectivists in psychology are without proper demonstration and reconsideration as long as they do not conserve the moment. It was suggested frequently therefore that "a talking machine should photograph the process" and that we should make systematic use of this machinery of personality recording.*

Psychodramatic procedure, being an offshoot of sociometry, has to be viewed and evaluated with sociometric methodology and experimentation as a background. Terms like *sociodynamic effect*, *social atom* and *tele* can appear esoteric and intuitive only to one who has not acquainted himself with the methods by which they have been explored.

* MORENO, J. L. *Application of the group method to classification*. New York: The National Committee on Prisons and Prison Labor, 1931, p. 16. (Current edition: *Group method and group psychotherapy*. New York: Beacon House, 1946, p. 16.)

BOOK REVIEWS

LANDIS, CARNEY, & BOLLES, M. MARJORIE. *Textbook of abnormal psychology*. New York: Macmillan, 1946. Pp. xii+576.

This is, for me, the best book of its kind. I am using it in both undergraduate and graduate courses in Abnormal Psychology. It is comprehensive and scholarly. It is a serious text, not the chatty type; though clearly written, it has the solid character of "an excellent technical summary."

It is an eclectic text. No single system of thought in psychiatry or psychology is preferred. Perhaps eclecticism is the only tenable position in the field at the moment, but such an approach is bound to seem desultory and, at times, confusing. I would not use the book as a text if I could not myself present a dynamic view of human personality to supplement the too-neutral picture here given. If an individual lecturer does not feel up to giving such a picture on his own account, perhaps he should use the Landis and Bolles text in conjunction with some other book which emphasizes a more dynamic outlook. A tremendous advantage of Landis and Bolles lies in *scope*, in the great amount of thoroughly metabolized information which is presented.

Landis and Bolles are much impressed by the twin data and by the data on inheritance of schizophrenia. I am still stubbornly unconvinced. Kallmann's study on *The Genetics of Schizophrenia* seems to me full of assumptions which need to be combed out by an expert in statistical design. Some ingenious skeptic with good training should attack this problem. If Kallmann's method should prove to be fallible we could spare ourselves much confusion by learning this at an early date. A single misguided but authoritative-seeming research can divert scientific activity from its proper goals for a considerable period. Despite such historical studies as Kallmann's, I shall remain in doubt about the inheritance of mental disorder until we have a science of learning applied to the problems of childhood and personal development.

The organization of the text is novel and useful. Psychoses, neuroses, and mental deficiencies are dealt with first. Then follow sections on heredity, cultural and personal development. Then the authors use an interesting device. They go back over the abnormal data from the standpoint of the psychologist, presenting chapters on disorders of sensation, speech, volition. In the case of speech disorders the picture would be like this: In the early chapters concerning psychoses, the slurred speech of the paretic or the labored vocalization of the depressed person will be duly noted. In the later chapter on speech, all speech deviations will be reviewed as a group and will be related to relevant psychological hypotheses and experimentation concerning speech development. These chapters give a specific content to the claim of the psy-

chologist for a special competence in dealing with the phenomena of mental disorder.

At the end of the text some omnibus chapters appear—on the rôle of the brain, legal psychiatry, and psychotherapy. The chapter on "The Law and Mental Abnormality" is a characteristic example of the thoroughness and general cultural interest of this book.

I am sure that my own research preoccupations will never permit me to undertake the tremendous task of writing a text in Abnormal Psychology. If they would, however, I would like to imitate Landis and Bolles in certain respects. I wish my imagined text would show as much of hard work, effort at neutrality of presentation, and broad base of facts as does theirs. I am sure I would be appalled at the conditionality of our information in this field. Most writers show an entire absence of methodological sophistication. Here and there one finds percentages cited, but the ideas of reliability of classification and validity of relationship among signs and events are woefully absent. The characteristic standards of the psychologist have never really been brought to bear on the field of Abnormal Psychology. The physicians who are the major reporters of evidence in this field inherit data from the pre-clinical biological sciences but rarely understand or use with independent force the *methods* of these sciences. The writer in this field has to be a sort of one-man reviewing stand, attaching his own estimates of probability to the data presented.

I am sure my text would make a greater effort toward a behavioral-psychoanalytic synthesis of life history materials. The lack of this in Landis and Bolles is fairly plain, especially in the clinical histories, which tend (as a result, in my opinion) to be dull and unilluminating. I would struggle to apply the "laws of learning" to the problem of child socialization, and to find in childhood the deviant habits which, if emphasized, might become traits in the later constellation of a mental disorder. I think this striving would drive me to write a more biased and contentious book than these authors have achieved, and this might well be a disservice to science.

I think it would be worthwhile to put in such a text a reminder to the teacher that audio-visual aids can perform a useful rôle in a course in Abnormal Psychology. Some of the better films on particular topics might be referred to. It might also be noted that anyone with a sound-recording gadget of modern fidelity can easily record characteristic interviews with mental patients, and thus achieve the tremendous value of the auditory stimuli in defining the neurosis or psychosis. However, I need hardly remind the reader that the bird in the hand provided by Landis and Bolles is worth a flock of these birds in the bush of my imagination.

JOHN DOLLARD.

Institute of Human Relations, Yale University.

HORSLEY, J. S. *Narco-analysis*. London: Oxford Univ. Press, 1946. Pp. vii+134.

"The purpose of this book is to describe a technique which (the author) devised in order to shorten the course of deep mental analysis and psychotherapy." Since psychoanalysis, because of its long duration and expense, is not practical in many situations, a method is provided utilizing drugs which overcome the initial resistance of the patient by "loosening his psychic tension" and thus facilitating the analysis of the patient's mental content.

After a brief survey of the history of narco-analysis, the author discusses the methods of hypnosis and indicates their basic shortcoming and the need for a method of inducing narcotic hypnosis. This is followed by a discussion of the chemical structure of the two different types of barbiturates—those characterized by anticonvulsant properties (containing the phenyl radical) and those characterized by hypnotic effects (containing only aliphatic substituent groups). He then proceeds to describe the application of sodium amytal to incommunicative schizophrenics, resulting in their return to normality for about two minutes during which they discussed their illness and future plans. This breaking through of the psychotic shell of the patient enables better contact later, so that study, diagnosis and even therapy can be introduced with the most recalcitrant of patients. This method is designated as the "drug-analytic" method in contrast with narco-analysis. The drug-analytic method consists essentially in giving intravenous injections of a barbiturate to produce deep narcosis, and in utilizing the patient's subsequent remarks as a basis for making better contact with him. The narco-analytic method is essentially psychotherapy conducted under light narcosis.

This chapter is followed by a section on narco-analysis in a mental hospital, out-patient clinic, and in wartime.

The particular psychological mechanism in which narco-analysis operates is then discussed. Abreaction and transference are both analyzed with reference to the proper role that narco-analysis may play in each. Illustrative case histories of schizophrenia, dream-analysis and simulation and malingering are given.

In his final chapter the author points out that in choosing a method for therapy the choice will depend upon the type of patient, time available, and inclination and experience of the physician. Hypno-analysis may be preferable in some cases, he says, since it does not depend upon the use of any drug or any physical means and the patient is fully conscious that improvement is due entirely to his own effort and not to some mysterious agency with consequent absence of insight. For inexperienced physicians, however, the drug-analytic method is to be preferred and the experienced therapist can use it to bring about short cuts in development of contact. Furthermore, the limited number of

cases to which hypno-analysis is applicable limits the usefulness of the method. Throughout the author stresses that the advantages of drug therapy itself would soon be lost if psycho-therapy were not introduced and utilized to its greatest extent.

JOSEPH ZUBIN.

Psychiatric Institute, Columbia University.

LOWREY, LAWSON G. *Psychiatry for social workers*. New York: Columbia Univ. Press, 1946. Pp. ix+317.

The modern social worker is in need of knowledge of psychiatric concepts and psychotherapy. What knowledge to serve what purposes? Out of his extensive experiences in the child guidance work, Dr. Lowrey is in a position to know. His *Psychiatry for Social Workers* represents an effort to satisfy this need in the light of his experiences. The book presupposes a familiarity with psychiatry by way of one course or the reading of a single text-book in psychiatry. The author does not intend the volume to serve as a book on the origin and nature of human behavior nor does he want the book to be a text in social psychiatry. What he does want is a volume that will serve social workers by furnishing them with knowledge of indicative symptoms of deviations in mental functioning with special reference to differential symptom pictures. The consequent content selected includes a description of the major types of psychoses, neuroses, mental deficiency, epilepsy, psychopathic personalities, and behavior disorders (due both to internal pathology as well as to reactive behavior disorders). These are presented under relatively static chapter captions. Illustrative are: "The Data of Psychiatry," "Clinical Combinations of Symptoms," "The Convulsive Disorders," "Psychopathic Personality," "Schizophrenia," and "Cyclothymic Psychoses." And these topics are relatively well considered for the purpose.

Many features that are not suggested by the captions are also included. Illustrative special features worthy of mention include: a telling use of admission rates to indicate the significance of the various mental diseases; a concise treatment of delinquency; a clinical description of the developmental stages without overdoing it; a consistently realistic and dynamic consideration of most problems; a distinct bias in favor of breaking through the unreal and nonexistent barriers of examination and treatment; a critical attitude towards the examination for examination's sake, history for history's sake, and other empty gestures; and a distinctly favorable response towards the application of the differential approach. By this approach therapists proceed in accord with needs revealed as the case itself unfolds, varying the tempo and sequence as indicated.

Psychological examination is dealt with in the chapter on "The Data of Psychiatry" and is mostly on the level of psychometric presentation,

as generally found in child guidance clinics directed by psychiatrists. The chief emphasis is on intelligence tests, performance tests, reading disability tests, and the Rorschach. Tests are barely mentioned and not too comprehensively considered. The Healy and Bronner manual, Terman's *Measurement of Intelligence*, Wechsler's *Measurement of Adult Intelligence*, two or three references on the Rorschach, and one reference on reading are all that psychologists would find on the strictly psychological phases of the examination. "The Data of Psychiatry" includes anamnesis, social examination, outline of a consideration of race differences, physical examination, as well as psychometric examinations.

All in all, problems considered are dealt with realistically as well as dynamically. In general, the book promises to satisfy the purpose for which it was written; namely, to serve as an intelligent guide for social workers.

H. MELTZER.

Psychological Service Center, St. Louis, Missouri.

ARTHUR, GRACE. *Tutoring as therapy*. New York: The Commonwealth Fund, 1946. Pp. ix+125.

In the current literature the role of the psychiatrist in solving school behavior problems has been given considerable stress. The role of the psychologist as a non-directive therapist has received recent attention. Dr. Arthur places a renewed emphasis on the therapeutic role of the psychologist as planner of remedial and individualized tutoring. Her book should be read by those interested in therapy techniques, child guidance programs or designs for curriculum flexibility to fit individual needs.

Dr. Arthur's principal theme is that the right kind of tutoring, provided at the right time, can work wonders for a child's adjustment at school, at home and in his social life. Examples of effective therapy are drawn from her experience as school psychologist and tutor, child guidance clinic psychologist, and director of a W.P.A. tutoring project. Her discussion of finding tutors, remedial methods, program rearranging, etc., is generously illustrated with individual cases summarized with affection and humor. Many of the cases are followed for the entire course of schooling. An evaluation is made not only of the immediate improvement in tutored subjects but of the subtle, long-term psychological adjustments as well.

Perhaps the chief defect of the book is that it paints too rosy a picture of tutoring therapy. Dr. Arthur emphasizes that good tutoring includes an extensive study of the child, a tutoring plan worked out by an experienced psychologist, a tutor carefully chosen with the child's particular psychological needs in mind, plus frequent retests and progress reports. As is customary successful cases are cited. A more ob-

jective tone could have been provided by including certain complete reports of children who did not respond adequately to the tutoring. That there were such children is obliquely suggested in the number of cautions which Dr. Arthur provides. For example: (1) the psychologist takes the sole responsibility for the tutoring plan, leaving the tutor free to side with the child; (2) tutoring programs should be tapered off rather than discontinued, etc.

Altogether this book provides a timely emphasis, certain good diagnostic insights, tutoring hints, and a wealth of illustrative cases.

ADELLA C. YOUTZ.

Teachers College, Columbia University.

WITMER, HELEN LELAND (Ed.). *Psychiatric interviews with children*. New York: Commonwealth Fund, 1946. Pp. x+443.

In her introductory remarks the editor states that "the aim of this book is to show various ways in which psychiatrists in child guidance clinics utilize the therapist-patient relationship for therapeutic ends." In each of ten case studies, the details of the successive interviews between child and psychiatrist are set down according to a uniform plan. The form of presentation is similar to that used by Carl Rogers in the case reported in his *Counseling and Psychotherapy* but these are not verbatim reports. No notes were taken while the interviews were in progress but presumably each was written up shortly after its conclusion.

The book is designed to illustrate the changing approach in handling children's behavior problems which has been taking place during the past ten or fifteen years. Before 1935, child guidance workers depended chiefly upon discovering and modifying those factors in the child's environment that seemed to be primarily responsible for his difficulties. The weakness of this approach lay in the fact that in so many cases it proved impossible to put the needed changes into effect.

As psychoanalytic methods for children became more completely developed and with the increasing popularity of play therapy the point of view shifted. Attention became focussed upon the child rather than upon his surroundings. Direct observation of his behavior by the psychiatrist himself was substituted for second-hand reports. An intimate and dynamic relationship between patient and therapist could then be created in which the child would feel free to express his fears and his hatreds, his anxieties and his desires without fear of rebuke or punishment. The establishment of such a relationship requires time and patience as well as skill. In the 10 examples given, the number of reported interviews ranged from 9 to 53.

The book is divided into three parts. Part I was written by the editor. It is entitled, *Introductory Comments* and contains three chapters. The first is largely historical, describing the conditions that led

up to the establishment of child guidance clinics and the changes in theory and procedure that have taken place since that time. The second chapter describes the kind of patients commonly accepted for treatment and the problems most often displayed. Chapter III is entitled, *Relationship: the Dynamic of Therapy*. It presents the modern theories underlying psychiatric work with children which, as indicated by the title, have one basic point in common. All demand the establishment of a feeling of security in the presence of the therapist such that the child becomes free to "be himself" without fear of censure or loss of "face." Different psychiatrists utilize this relationship in various ways but all recognize it to be essential for the success of the treatment.

Part II which constitutes the main part of the book begins with an introductory chapter by the editor describing the criteria used in the selection of cases. These were as follows: (a) the psychiatrists making the reports should be known to the editor as having had thorough training in child therapy and extensive experience in child guidance work, (b) only those cases should be included in which the treatment proved to be at least reasonably successful, (c) the cases selected should be fairly representative of the type commonly brought to child guidance clinics and should not include any unusual or bizarre examples. As far as possible they should represent a wide range of ages, problems and home backgrounds and (d) no foster children should be included; all must be living with their own parents.

The case reports which follow were selected by the editor in cooperation with the psychiatrists who prepared them. The following psychiatrists contributed to the study: Frederick M. Allen, Phyllis Blanchard (two cases), Lydia N. G. Dawes, Hyman S. Lippman, Martha W. MacDonald, H. B. Moyle (two cases), Beata Rank, and Robert A. Young. The general plan of presenting the reports was outlined by the editor but the actual case material was written by the psychiatrists. For each case there is given first a brief paragraph describing the child and his environment, the type of problem presented and the manner in which the psychiatrist utilized his relationship with the child in order to work out a solution. The editor then elaborates upon this material in order to show her reasons for selecting this particular case for inclusion in the study. This is followed by the social worker's report of her initial findings after which come the psychiatrist's reports of the successive interviews. The reports are interlarded with copious footnotes in which the psychiatrist gives his reasons for directing the interview as he did, and presents his interpretations of the child's reactions. The majority of the case reports are interrupted midway in order to give an account of the social worker's progress in working with parents. Each report concludes with a brief account of follow-up work over a varying period of time.

In Part III entitled *Comments in Conclusion* Dr. Witmer reiterates

her earlier warning that this book should not be regarded as a report of an experimental study nor should the accounts of the psychiatric interviews be looked upon as models to be followed by rule-of-thumb. The cases selected are not representative in the statistical sense of the word but only in the popular sense in that the children and their problems are similar to those of many other cases brought to child guidance clinics. The fact that only cases with reasonably successful outcome were chosen obviously means that only a limited part of the usual intake of a child guidance clinic would be represented at all. With these limitations in mind, Witmer then discusses certain points of similarity and dissimilarity in the way the various cases were handled, and stresses the fact that superficial resemblance does not always mean basic-similarity. Differences in treatment arise in part from differences in the concepts held by various therapists as to how their relationship to the child can best be utilized; in part from differences in the children and in their problems. Always, however, the effort must be mutual. The psychiatrist does not attempt to solve the child's problems for him; he only helps him to find a solution for himself.

The book is of interest to psychologists for two reasons. In the first place the boundary between psychiatry and clinical psychology is far from clear-cut. The aims are similar and the procedures overlap. Clinical psychologists will therefore find in it many useful suggestions and will be stimulated by its argument, whether or not they are in agreement with the point of view expressed. Psychologists who are not clinicians will also enjoy the editor's clear and scholarly exposition of principles and theories with which all should have some familiarity even though many will hesitate to accept them without qualification. Although statisticians may cavil at the tacit assumption that the reported improvement in the attitudes and behavior of the children which ensued upon treatment was due to the treatment and not to other factors taking place simultaneously, the book remains a notable contribution within its field since it is the first detailed account of long continued therapeutic treatment of children by means of interviews in which the successive steps and the reasons for taking them have been presented in an orderly and systematic manner.

FLORENCE L. GOODENOUGH.

University of Minnesota.

WOLFF, WERNER. *Personality of the preschool child*. New York: Grune & Stratton, 1946. Pp. xvi+341.

This book is, as the author remarks, a result of "the urge to attempt an approach to the hidden forces and motivations of child behavior." Throughout the publication the author emphasizes two themes, that child behavior is a meaningful and interpretable product of the dynamics of the child's personality, but that the interpretation must proceed from

a recognition of the essential differences between the child and the adult. Even when the child's behavior is "sensible" in adult terms, it is dangerous to assume that it is the result of adult-like personality dynamics.

Dr. Wolff divides the book into three sections. The first is a discussion of the characteristics of children based upon observations of free play, supported by references to the findings of other investigators. The second section is a report of a number of experimental procedures, for the most part interpretation and discussion of the pictures drawn by children under experimental conditions. In a final section the general theoretical conclusions are summarized.

It is not possible in this review even to list the specific experiments performed, or to report the details of the author's theories. He continually emphasizes the fact that the child does not display discrete behavior which is a product of one aspect of his personality, independent of the rest. Where the adult is able to perceive a situation or perform a task relatively uninfluenced by random thoughts and extraneous stimuli the child's perception and behavior is not so differentiated. This characteristic plus the fact that the child's behavior omits many of the connections between ideas makes for difficulty in interpretation.

To some extent the book itself shows the same lack of differentiation. It is extremely difficult to describe behavior which is so diffuse and so greatly influenced by numerous factors and still to keep the description itself clearly organized. Another characteristic of the book which makes it difficult to read is the small amount of space given to some topics. The text is most readable and most convincing in the long descriptions and discussions of the drawings of an individual child; it is least satisfactory when the author tries to summarize his opinions on some complex topic in a few categorical sentences.

Though difficult to read, the book is worth the effort; this reader came away with the opinion that Dr. Wolff has, by his experimental techniques and demonstration of interpretative methods, made an important contribution to the psychology of the preschool child.

There is a bibliography of 628 titles and an excellent index.

ALFRED L. BALDWIN.

Fels Research Institute.

RADKE, MARIAN J. *The relation of parental authority to children's behavior and attitudes.* Univ. of Minnesota, Institute of Child Welfare Monogr. No. 22. Minneapolis: Univ. of Minnesota Press, 1946. Pp. x+123.

In recent years psychologists have been attributing increasing importance to the cultural milieu as a formative factor in various aspects of the personality. It has been said that family constellations, methods

of discipline, factors of security or insecurity, and so on, influence a person's attitudes and the adequacy of his emotional adjustments. Radke, in this book, reviews the literature—clinical, semi-experimental and experimental—of the initial attempts to evaluate accurately the effect of cultural influences in the area of parental-child relationships, and in the use of projective techniques in studying personality. She then goes on to report an experiment of her own, on 43 four-year-old nursery school children and their parents. She has selected for her study one small area of the culture, that of parental authority and discipline, and a small fairly homogeneous upper-middle-class population.

From the review of the literature, the researches in this field are shown to be thus far tentative, inconclusive, and in some respects contradictory. Radke's experiment is devoted in part to an evaluation of the methods used. She employs interviews with both the parents and the children, teachers' ratings, controlled behavior situations, and projective techniques (in the form of doll play and responses to pictures). The various kinds of data serve to check each other, and to assist in the evaluation of each, as well as to indicate the influence of some of the cultural influences observed.

It was found that the children reported accurately on many of the objective facts of parental behavior and discipline, but that their interpretations were limited.

Several factors interfere with any definitive demonstration of strong relationships. For one thing, as Radke points out, relatively small differences are to be expected when the population studied is so homogeneous in many aspects of its culture. One important point for consideration in further research in this field is the repeated finding that the children's understanding and interpretation of the home situations, and of specific disciplinary measures, are often quite different from the interpretations and intentions of the adult experimenters and disciplinarians. Care needs to be taken in evaluating the situation from the child's point of view if we wish to understand the effect of his environment upon him. Even with these limitations, the data show relations between the degree of parental restrictiveness and child behavior. These differences are shown to be similar in many respects to those of other experimenters using similar classifications of adult authority over children. There are other interesting findings on sex differences in attitudes and disciplinary measures.

The experiments reported here are carefully done, with due regard for scientific procedures and controls in investigating characteristics which until recently have been based largely on clinical experience and intuition. They indicate important considerations for future research in this area, and for parent-education in aspects of child discipline.

NANCY BAYLEY.

University of California.

MAURER, KATHERINE M. *Intellectual status at maturity as a criterion for selecting items in preschool tests*. Univ. of Minnesota, Institute of Child Welfare Monogr., No. 21. Minneapolis: Univ. of Minnesota Press, 1946. Pp. ix+166. \$2.50.

The problem attacked in this monograph is that of selecting the subtests of the Minnesota Preschool Scale (both forms) which are most predictive of intelligence at maturity as measured by the Wells revision of the Army Alpha. Of 1091 cases on file who had been tested prior to age six, 226 were located and tested at ages 16 to 22. A grand total of about 700 biserial correlations, between Alpha scores and the Preschool Scale items, were computed on 180 cases subdivided into nine age groups according to age at time of their preschool test (N 's varied from 9 to 33). The records of 46 cases (every fifth) were held for "validating" the "predictive" scale made up of the subtests selected on the basis of the item analysis.

Two points lead the reviewer to question the accuracy of the statistical analysis. With sizable standard errors (up to .30), it is very doubtful that the 400 r 's under .40 (200 under .20) would all be on the positive side of zero; yet no negative biserial values are reported. About one-tenth of the 700 given standard errors, for biserials, are erroneously low.

On the basis of the item biserial r 's, 17 of the 29 subtests of each form were retained as "predictive." These 17 represent about 75% of the 224 scoring items per form. The preschool test blanks of the 46 validating cases were rescored on these selected items, and the correlations of Alpha with selected tests were compared with the correlations yielded by all the preschool subtests. With r 's based on N 's of 3 to 8 per age group, the results are inconclusive: it cannot be said that the scores based on selected tests are any more predictive than scores based on all the tests. This lack of conclusiveness also holds for the findings based on a supplementary validation group with Stanford-Binet IQ's at ages 11 to 15 taken as measures of near terminal intelligence.

A better procedure would have been to choose on the basis of the item analysis the best half of the items as predictive, with the remaining half as nonpredictive, then to compare the validity of the predictive versus that of the nonpredictive tests, instead of the predictive versus the total. The difficulty of eliminating the age factor so as to have a validity coefficient based on all 46 cases could be overcome by rescored a sufficient number of typical Minnesota Preschool Test blanks as a basis for standard scores, at each age level, for the predictive items. These standard scores could then be correlated with later Alpha standing, also expressed as standard scores by age level.

This monograph represents an enormous amount of work on a very important, and likely fruitful, line of research. Far more cases will be required for checking the reasonable hypothesis that preschool scales

contain "dead wood" as far as the prediction of terminal intelligence is concerned.

QUINN McNEMAR.

Stanford University.

FORD, MARY. *The application of the Rorschach test to young children.* Univ. of Minnesota, Institute of Child Welfare Monogr., No. 23. Minneapolis: Univ. of Minnesota Press, 1946. Pp. xii+114. \$2.00.

Dr. Ford's study was initiated in 1936; many of its goals and methods were set at that time. In the face of the enormous proliferation of Rorschach interest, Rorschach studies and Rorschach users (and misusers) in the past decade, what would once have been of pioneering significance is now more in the nature of confirmation, elaboration (or, occasionally, refutation) of studies published in the interim.

The first problem Dr. Ford set herself, to refine and simplify Rorschach administration, was not resolved by research, but rather by preliminary experience. Most notable modifications are the use of a trial blot à la Hertz, a prohibition against turning the cards, and a certain amount of "prodding" for further responses. Undoubtedly there will be disagreement regarding the need for such changes. Although she appears to have been unusually successful in carrying through her inquiries, Ford, like others who have worked with young children, suggests the desirability of a brief inquiry at the time each card is shown. The study leaves open the question of the lower age limit for the administration of the Rorschach, inasmuch as no children under three were tested. Dr. Ford confirms and further develops Klopfer and Margulies' description, in 1941, of the early "stages" of Rorschach response.

Her second problem was the study of "variations in the test determinants" (scoring categories) and their relation to sex and age. No normative value is claimed because the 123 subjects "represent a highly selected group both in intelligence and in socio-economic status" (p. 39). Existing adult and adolescent norms were used. This section cannot but remind the reader of the regrettable lack of final agreement on scoring nomenclature. In the few instances where the present study is at variance with the Klopfer-Margulies' findings this appears to have been a determining factor. (For example, including among whole responses the "cut-off whole," by Klopfer and Margulies but apparently not by Ford; or again, "One possible explanation for the low incidence of C responses in . . . the present study is the fact that Cn responses were not included in C responses" (p. 47) as is customary and was the case in the other report.) Certain other discrepancies raise the more serious question of disagreement on scoring. For example, the author points to the higher incidence of M in her study in spite of the fact that "all the investigators including the writer, use Rorschach's instructions in scoring M re-

sponses" (p. 54). Yet in the three sample protocols several instances appear to the reviewer in which movement was scored even though not "human or human-like." (E.g., "lots of bugs crawling around.") The reviewer would also quarrel with the scoring of some of the color responses. In general, sex differences appear unimportant. However, Ford shows that many important scoring categories bear a clear-cut relationship to chronological age between 3 and 8. These findings, along with others in the literature, point to the need of establishing norms perhaps for each year-level under 7 or 8.

An attempt was made to validate the relationship between intelligence and certain specific determinants suggested by Rorschach. Satisfactory correlations were found for most of these but not for W and %A. These exceptions "make sense" in view of the fact that children's whole responses are often crude and undifferentiated, unlike the more structured W's of adults, and in view of the fact that animal concepts in children's thinking do not have the banality that they do for adults.

A third contribution of the study was in establishing test-retest reliabilities for the specific determinants of the Rorschach with young children. Most of the coefficients are of the order of $\pm .7$. As Dr. Ford points out:

The evidence of this investigation suggests that with further refinement of procedure and scoring, the reliabilities of the various test determinants may be expected to approximate those of other standard tests for young children. Even at the present time, under the conditions set up in this study, the reliabilities of most of the determinants are fairly satisfactory, especially when one considers that the interpretation of test results does not depend upon any one determinant but upon the balance and interrelationship between determinants.

Her fourth major aim, the study of the "validity of some of the determinants of personality," Dr. Ford undertakes with great modesty, but it is difficult to evaluate because data so meticulously supplied in other sections of the monograph are inadequate here. As one example, in evaluating emotional adjustment on the Olson scale we are told that the subjects were "divided into three groups on the basis of emotional adjustment as indicated by the Rorschach Test" (p. 82), but we do not know how this was done. A new scoring concept of *organizational links* (OL) is tantalizing because it is always referred to allusively. Defined in the glossary as "The number of verbalized, logical connections within a response or within a record," this would appear to be a valuable indication of complexity of response. However, the sample protocols make it difficult to determine whether one counts links or items. "A man walking a fence" receives a score of 2 OL; "a partly burned tree trunk" is also scored 2OL; while "a balanced snake" is scored 1 OL. For the clinician there is some disappointment in the mass treatment of data and the paucity of individual records or even scoring lists. However, he

will find much of interest in the discussion of administration of the Rorschach and on the frequencies of locations, determinants and content. The student of personality development will discover certain interesting findings—and rather more interesting suggestions—for the use of the Rorschach to study development in general, and the development of the process of logical thinking in particular.

L. JOSEPH STONE.

Vassar College.

KABACK, GOLDIE RUTH. *Vocational personalities: an application of the Rorschach Group Method*. Teachers College, Columbia Univ. Contr. to Educ., No. 924. New York: Bureau of Publications, Teachers College, Columbia Univ., 1946. Pp. x+116.

The problem of the relationship between personality and vocational choice has been of long standing interest. In the above research an aspect of this problem is studied by means of the Group Rorschach Method. The vocations investigated are primarily Pharmacy and Accountancy although comparisons are also introduced with some other groups for which published data are available. The author selected 75 subjects in each of four groups, consisting of practicing pharmacists, practicing accountants, and students preparing for the two professions. In addition to taking the Group Rorschach Test, each subject answered a brief questionnaire pertaining mainly to vocational aspirations, training and experience, and to leisure time activities. Practitioners averaged 37 years in age. Pharmacy students averaged 20 years and accountancy students 18 years.

The bulk of the monograph is devoted to a meticulous statistical analysis of inter-group differences based on most of the large variety of scorable Rorschach factors that have been recognized to date. The statistical analysis is conscientiously performed and well presented but the scope of the study does not go much beyond statistically defined problems. A limitation is also placed on the data by dealing with the Rorschach factors in a purely quantitative fashion. The interpretations of the data are thus based only on the sheer number or the proportions of various significant items. The omission of qualitative considerations can at times be disastrous to an intensive character analysis.

A number of statistically significant differences are obtained between pharmacists and accountants and between the corresponding student groups. The accountants have a higher level of intelligence (as judged by a number of different items) and a greater degree of caution. The pharmacists have a greater degree of anxiety. There is, however, a considerable amount of overlap between the distributions of most of the Rorschach factors. The author concludes:

The most notable feature of the study is that both professions attract people of diverse personality characteristics (p. 83).

Each profession appears to attract persons with varying degrees of emotional stability, introversiveness, extratensiveness, intellectual level, etc. (p. 82).

A fuller understanding of the relationship between personality and vocational choice requires a broader investigation than the one reported here. Vocational choice is related to the prestige, economic reward, economic security afforded by the vocation as well as to the skills demanded by it. The prestige may vary with the social stratum from which an individual comes, as may the standards which determine adequate reward and adequate security. An analysis of what the appeals of given occupations are could be the basis of more sharpened hypotheses as to which personality factors to investigate. An attempt to obtain such information, in the present study, through a questionnaire was not very successful. When asked the reason for their choice of profession, over 60 percent of the 300 subjects either did not answer this question, or said they selected the profession because they liked it or were interested in it. Intensive interviews of a few subjects might have been more fruitful, perhaps yielding a picture of the relationship between self evaluation and occupational choice. Such findings as the greater anxiety among pharmacists could be clarified if there were greater understanding of the meaning of the profession as a way of life to the people in it.

The discovery of a high degree of overlap between the pharmacists and the accountants, despite the wide differences in the nature of these professions, focusses attention on the limitations of a personality approach to vocational choice in present day society.

As the author indicates, the present study cannot be taken as a final statement. She points to the need of a more diverse sample than was used and to the need for further experimentation with the Group Rorschach Method to clarify interpretations which were carried over bodily from the individual test. We would stress the value in future research of an intensive approach which would not limit personality description to the special characteristics of a single test and which would investigate motivational and social determinants as well as personality structure.

MAX HERTZMAN.

College of the City of New York.

LAWTON, GEORGE. *Aging successfully*. New York: Columbia Univ. Press, 1946. Pp. xiv+266.

DE GRUCHY, CLARE. *Creative old age*. San Francisco: Old Age Counseling Center, 1946. Pp. 143.

The normal "problem group" in our society is the later-maturity or old-age group; and this group is increasing at a rate which compels attention, if not alarm. Thus, "in 1950 the age group forty-five and over will make up 30 percent of the population" (Lawton, p. 257). By 1980,

the number of persons over 45 will have risen to about 40 percent of the total population.* If psychology were more alert, the science of *geronchology*—if one may coin such a term for the psychology of aging—would occupy a spot much closer the center of the psychological stage than it now does.

Both the books by Lawton and de Gruchy are "popular," in the sense that they are well-written, and aim to reach an audience that is more interested in practical application than in scientific proof. The book by de Gruchy is divided into two parts. The first part presents case-histories of seven individuals counselled by the San Francisco Old Age Counselling Center (founded by Dr. Lillian J. Martin); the second part describes principally the "SOA Farm"† and the development of various avocations by a number of counselees. As R. G. Kuhlén states in his excellent Foreword to the book, the keynote of the San Francisco Counselling Center is the stimulation of goal-directed activity on the part of old people. Neither de Gruchy nor Lawton has much faith in the adequacy of merely recreational or "chronicidal" (time-killing) activity. "What is needed instead is the self-fulfillment of the individual, through securing for him an opportunity for growth in his preferential field of interest" (de Gruchy, p. 110). De Gruchy cites a striking example of the inadequacy of the merely recreational approach:

In Long Beach, California, perhaps the town most fully populated by old men and women, the municipality has recognized some needs for the old. Many forms of suitable play have been established. Bowling and croquet lawns abound. There are open-air shelters for card playing, billiard tables, and the like. But watch the people who use them. They are killing time, perhaps even mildly amusing themselves; but they are as dour-looking, as slowed-down, as limp and lifeless as any other old age group found in less privileged locations (p. 88).

Lawton's book is more comprehensive than de Gruchy's, aiming to provide, in an informal or conversational manner, guidance "on the most effective way of making the typical adjustments called for by later maturity" (p. 236). The word "typical" in this quotation jars the present reviewer, because it seems to imply some lack of recognition of individual differences. Both Lawton and de Gruchy give their enthusiastic, general approval to an "activity" program: namely, the development of personally significant goals, purposeful activity, and future-mindedness (this advice to senescents sounds strangely like common advice to adolescents). Very likely such a prescription fits the needs of most counsellors' clients. But what about the underweight, cardiac neurasthenic who might benefit from more rest and less purpose; or the

* DUBLIN, L. I. Longevity in retrospect and in prospect. In E. V. Cowdry (Ed.), *Problems of Ageing*. Baltimore: Williams and Wilkins, 1942.

† The letters SOA stand for "Salvaging Old Age."

over-active oldster with plenty of purpose, who fails to recognize his own limitations and gets into everyone's hair in the process? "Calm down" is occasionally as good advice as "get up!" Purposeful activity may well be essential to self-respect and social approval; nevertheless, the realistic adaptation to old age also demands the development of a certain degree of relaxation, of satisfied tolerance for idleness and diminished functioning. The reviewer feels that Lawton's and de Gruchy's energetic activity-prescription must be applied with discrimination—most likely the very kind of discrimination which these consultants apply in their practice, but which they have failed to reflect in their books.

Both Lawton and de Gruchy write with infectious enthusiasm and zeal. Both feel sure that their methods are highly successful. For the San Francisco Old Age Counselling Center, de Gruchy cites an actual figure: 9 percent failure, 91 percent success. The reviewer is surprised that de Gruchy does not present separate figures for men and for women; since the adjustment problems of most men (who pass from a paid occupation to unflattering and unremunerative retirement) seem obviously greater than the problems of most women (who simply continue as housewives). In any event, the figure of 91 percent success is very impressive; and as Kuhlen points out in his Foreword to de Gruchy's book, "The danger of interpreting a transitory 'pepping-up' as a significant change in personality and purpose has been avoided by refusal to evaluate success until at least two years have elapsed" (p. ii). Nevertheless, there are still grounds for skepticism. Thus, the reviewer doubts that either de Gruchy's or Lawton's samples represent a cross-section of the aged population with respect to health, financial resources, general level of ability, special abilities, and general level of social skills. Furthermore, as Kuhlen points out, the success of old-age counsellors may be "partially explained by the probability that the old person who bestirs himself sufficiently to seek aid and voluntarily to undertake a program of rehabilitation already has a latent motivation that can be capitalized" (p. ii). A counter-hypothesis is that only those with the more serious problems are likely to seek a counsellor; but the reviewer, for one, would give Kuhlen's suggestion the greater weight. In the case of Lawton's clients, the reviewer gained the impression that these were typically in the younger old-age group (below 65, rather than above)—which again seems a factor favorable to successful adjustment or rehabilitation.

Lawton's chapter, "In youth prepare for age," makes a point which deserves far greater emphasis than psychologists or educators have given it. Adult education seems limited in this respect, because—

Most men and women get so concerned about work that in the feverish desire to rise in the world they throw out the ballast of rest and play. They become creatures of one function, struggling unceasingly to meet competition, and money becomes a goal for living rather than a means to it (p. 102).

Most self-help books are guilty of the fallacy of voluntarism. While there are undertones or overtones of voluntarism in both Lawton and de Gruchy, these authors should be complimented for preserving, for the most part, a sensible means-end approach. On the other hand, both authors seem over-enthusiastic or over-optimistic regarding the possibility of surmounting the problems and handicaps of old age. Enthusiasm and optimism, however, are part of the therapeutic process, and (if not overdone) are probably justifiable in popular works.

The literary quality of the two books calls for some special recognition. Especially in Lawton's volume, there are some passages which remind one of William James. Three examples must suffice:

... the life which lasts longer and is progressively emptier—this is a peculiarly modern form of torture.

Caution and comfort become rallying cries as the old man races off to do battle without leaving his easy chair.

Devotion to "remembrance of things past" is for writers legal tender of high purchasing value. But for an older person, it represents inflated currency; it buys nothing and if he tries to live on it he will starve. Only an active creative present has true purchasing power for an older person.

Finally, it may be pointed out that both Lawton's and de Gruchy's books adopt the individualistic or psychological, rather than the sociological, approach. Both are books of counsel, not social reform. Lawton seems to recognize—but inadequately, we think—that the economical and practical solution of many of the problems of our "senior citizens" requires organized, governmental action. For de Gruchy, society and government apparently exist only vaguely in the background. Such a highly individualistic outlook runs counter to the view (held by virtually all serious students in this field) that "not all of the answers to the problems of gerontology are to be found within the individual."*

H. S. CONRAD.

College Entrance Examination Board.

MICHOTTE, A. *La perception de la causalité*. Études de Psychologie, Vol. VI. Paris: Vrin, 1946. Pp. viii+296.

The investigation of complex perceptual processes progresses very slowly. Rubin's book on figure and ground and Katz's book on the phenomenal appearance of colors presented two important steps forward. Now there appears a book on the perception of causality by the Belgian psychologist Michotte which, like these older books, seems to open up a new field. The book is largely concerned with the answer to the question of what the stimulus configurations are which arouse in the

* OVERHOLSER, W. Orientation. In *Mental Health in Later Maturity*. Supplement No. 168 to the Public Health Reports. Washington, D. C.: Government Printing Office, 1943.

observer the impression that one thing is pushing or pulling another.

A very ingenious method is used for producing the variation in the stimulus patterns. Lines of different shapes are drawn on a disk turning behind a screen with a horizontal slit. A ring on the disk produces a stationary spot, a spiral produces a moving spot. In one of the basic experiments, a spot *A* appears in the middle of the slit. Another spot, *B*, appears from the left and moves toward the right till it touches *A* which has so far remained stationary; then *B* stands still and *A* moves to the right. This arrangement produces the phenomenon of *B* pushing *A*, the motion of *A* seems to be caused by the motion of *B*. The set-up allows changes in different variables: the time interval between the contact of *A* and *B* and the start of the motion of *A*, the speeds of *A* and of *B* and their relation to each other, etc.

On the basis of many experiments Michotte concludes that the phenomenon "lancement" (pushing by impact), to which one third of the book is devoted, is a unified perceptual form like a shape and that it arises under specific stimulus conditions. In the same way the phenomenon "entraînement" (pushing during contact) is investigated. Michotte adds a further theoretical analysis which can only be indicated in a brief review. He states that the movement of the agent seems to extend over the passive body, and that a unit formation between active and passive body plays an important role. Besides these fundamental phenomena the perception of the causation of qualitative changes and of animal locomotion are discussed, and the bearing of these findings on the interpretation of different philosophical theories of causality is considered.

The book contains a wealth of stimulating experimentation and theory in a practically virgin field. The only question that arises in the reviewer's mind is whether some results are not conditioned by the limited possibilities of patterns which the method of the disk allows.

FRITZ HEIDER.

University of Kansas.

DELGADO, HONORIO. *La personalidad y el carácter*. (2nd Ed.) Lima: Editorial Lumen, 1946. Pp. 207.

The author of this brief but comprehensive *précis* of personality and character has every reason to be gratified over the need for a second edition. Very likely his own influence as a distinguished psychologist and as editor of an important psychiatric journal had much to do with producing a continued demand for the book, which deserves to become as widely known in North America as it evidently is in South America. Added to the first idea of acquainting readers with the fruits of Dilthey's tradition in the field of personality, is the hope expressed in the new preface that irresponsible forms of popular characterology may be over-

come through the kind of balanced scientific treatment that Delgado offers.

Apart from the general theory of personality, covering structure, genesis, and dynamics, the plan of the book deals at length with some six typologies and with characterology as such. The limitations of all typological approaches are stressed, and there is a regular insistence on the cohesive nature of personality as well as a consistent demand for what Allport calls the idiographic approach. The typologies presented include temperament theory, and the schemes of Kretschmer, Jaensch, Pfahler, Jung, and Spranger. Such specialized studies as that of twins reared apart are reviewed in order to evaluate environmental factors. Freud's contributions are not regarded as foundational.

In the United States, books on personality frequently give short shrift to the topic of character. On the contrary, Delgado considers character at virtually every point in his discussion of personality, presenting the one as what amounts to an aspect of the other. To him character "in the strict sense, is personality manifested." The entire general theory of personality and character, as the reader is notified in the preface, takes its departure from a conception of spiritual integration rather than from biological considerations, and is frankly teleological in its implications. The fact that this will not interest certain students of personality in this country is perhaps a good reason in the long run for their becoming acquainted with Delgado's approach, and for observing its amplified scope as compared with narrowly operational investigations.

The changes between editions are all minor in nature, involving incidental corrections and a few small additions. The book contains well-arranged indexes and a useful bibliography. Its format is the tasteful and readable typography of the first edition.

American International College.

HOWARD DAVIS SPOERL.

LINK, HENRY C. *The rediscovery of morals, with special reference to race and class conflict*. New York: Dutton, 1947. Pp. 223.

The thesis of Max Weber was that a symbiotic relationship existed between Protestantism and Capitalism, so that they thrived together. Dubious as Weber's hypothesis may have been, if confirmatory illustration were wanted it could be found in this book by Dr. Link. The substance of the moral theory presented is that the freedom and dignity of man rest upon the institution of private property—or, at least, that threats to private property are the greatest threats to freedom.

The two moral sores festering in modern life are class conflict and race conflict. These are more or less equated as parallel problems, because both arise out of failure to respect the worth of the individual, and in both cases the cause of conflict is falsely attributed to economic causes

instead of moral ones. A generally wholesome and humane position on the race issue is used alongside a violently anti-union and anti-socialization position on the economic issue. On the one hand, the right of the Negro to a job parallels the right of the employer to make all the money he can. The moral law blesses both. On the other hand, the danger of a Fair Employment Practice Committee is that it might increase racial tension as the Wagner National Labor Relations Act has increased class strife.

Despite a certain show of evidence in references to a 72-item bibliography, including the results of public opinion polls and tests of personality quotients, the book is essentially an essay in defense of private enterprise, using the sanctions of decency, morals, and religion to bolster an argument. The tactics are those of the debater and propagandist, not those of the scientist, philosopher, or theologian.

ERNEST R. HILGARD.

Stanford University.

BOOKS AND MATERIALS RECEIVED

CASSIDY, ROSALIND, AND KOZMAN, HILDA CLUTE. *Counseling girls in a changing society*. New York: McGraw-Hill, 1947. Pp. xix+441. \$3.75.

DAVIS, HALLOWELL, *et al.* *Hearing aids: an experimental study of design objectives*. Cambridge: Harvard Univ. Press, 1947. Pp. viii+197.

DENNIS, WAYNE, *et al.* *Current trends in psychology*. Pittsburgh: Univ. of Pittsburgh Press, 1947. Pp. vii+225. \$3.50.

ESTABROOKS, G. H. *Spiritism*. New York: E. P. Dutton, 1947. Pp. 254. \$3.00.

GEMELLI, AGOSTINO, AND ZUNINI, GIORGIO. *Introduzione alla psicologia*. Pubblicazioni dell' Università Cattolica del Sacro Cuore, Nuova Serie, Vol. XX. Milano: Società Editrice "Vita e Pensiero," 1947. Pp. xiii+44. Lire 900.

GESELL, ARNOLD, AND AMATRUDA, CATHERINE S. *Developmental diagnosis: normal and abnormal child development*. (2nd Ed. Revised.) New York: Hoeber 1947. Pp. xvi+496. \$7.50.

GIBSON, JAMES J. (ED) *Motion picture testing and research*. Army Air Forces Aviation Psychology Program, Research Reports. Report No. 7. Washington: U. S. Government Printing Office, 1947. Pp. xi+267. \$1.50.

HARMS, ERNEST (ED.). *Handbook of child guidance*. New York: Child Care Publications, 1947. Pp. 751.

HOBBS, NICHOLAS (ED.). *Psychological research on flexible gunnery*

training. Army Air Forces Aviation Psychology Program, Research Reports. Report No. 7. Washington: U. S. Government Printing Office, 1947. Pp. xi+267. \$1.50.

JERSILD, ARTHUR T. *Child Psychology*. (3rd Ed.) New York: Prentice-Hall, 1947. Pp. xi+623. \$5.00.

JOAD, C. E. M. *How our minds work*. New York: The Philosophical Library, 1947. Pp. 113. \$6.00.

KARPMAN, BEN. *Case studies in the psychopathology of crime*. Vol. 2 (2nd Ed., Revised). Washington: Medical Science Press, 1947. Pp. xxix+738.

KÖHLER, WOLFGANG. *Gestalt psychology: an introduction to new concepts in modern psychology*. (Rev. Ed.) New York: Liveright, 1947. Pp. 369.

KUNZ, HANS. *Die anthropologische bedeutung der phantasie*. I. Teil: *Die psychologische analyse und theorie der phantasie*. Pp. xii+275. II. Teil: *Die anthropologische bedeutung der phantasie und ihre voraussetzungen*. Pp. 311. Basel: Verlag für Recht und Gesellschaft, AG. 1946. (American Representative: A. J. Phiebig, 545 Fifth Ave., New York 17, N.Y.)

MERRILL, MAUD A. *Problems of child delinquency*. Boston: Houghton Mifflin, 1947. Pp. xxiii+403. \$3.50.

MURPHY, GARDNER. *Personality: a biosocial approach to origins and structure*. New York: Harper, 1947. Pp. xii+999. \$5.00.

NAUMBURG, MARGARET. *Studies of the "free" art expression of behavior problem children and adolescents as a means of diagnosis and therapy*. Nervous and Mental Disease Monographs, No. 71. New York: Coolidge Foundation, 1947. Pp. xi+225. \$5.50.

NIELSON, J. M., AND THOMPSON, GEORGE N. *The engrammes of psychiatry*. Springfield, Ill.: Charles C Thomas, 1947. Pp. xix+509. \$6.75.

PEATMAN, JOHN GRAY. *Descriptive and sampling statistics*. New York: Harper, 1947. Pp. xviii+577.

PONDER, ERIC, *et al.* *Some aspects of red cell production and destruction*. Annals of the New York Academy of Science, 1947, Vol. XLVIII, Art. 7. Pp. 577-704.

SHERMAN, MANDEL. *Psychology for nurses*. New York: Longmans, Green, 1947. Pp. xiii+237. \$2.75.

SUMPTION, MERLE R. AND BEEM, HARLAN D. *A guide to school reorganization in Illinois*. Circular No. 59, Bureau of Educational Research, College of Education, Univ. of Illinois. Urbana: Univ. of Illinois, 1947. Pp. 52.

TIFFIN, JOSEPH. *Industrial psychology*. (2nd Ed.) New York: Prentice-Hall, 1947. Pp. xxi+553. \$4.00.

WICKERT, FREDERIC. *Psychological research on problems of redistribution*. Army Air Forces Aviation Psychology Program, Research Reports. Report No. 14. Washington: U. S. Government Printing Office, 1947. Pp. vii+298. \$1.25.

WILLIAMS, R. M. *The reduction of intergroup tensions: a survey of research on problems of ethnic, racial and religious group relations*. Social Science Research Council Bull. 57. New York: Social Science Research Council, 1947. Pp. xi+153. \$1.75.

ZEISEL, HANS. *Say it with figures: a manual for statisticians in the field of consumer and opinion research*. New York: Harper, 1947. Pp. xvii+250. \$3.00.

ZUNINI, GIORGIO. *Animali e uomo: visti da uno psicologo*. Milano: Societa Editrice "Vita e Pensiero," 1947. Pp. xi+293. Lire 300.

Bibliographie der philosophischen, psychologischen und pädagogischen literatur in der Schweiz 1941-1944. Jahrbuch der Schweizerischen Philosophischen Gesellschaft. Basel: Verlag für Recht und Gesellschaft AG, 1945. Pp. 218-278. (American Representative: A. J. Phiebig, 545 Fifth Ave., New York 17, N. Y.)

Boletin del Instituto Psicopedagogico Nacional. Vol. VI, No. 1. Lima, Peru: Solicitamos Canjes, Apartado No. 636, 1947.

1947 Achievement testing program in independent schools and supplementary studies. Educational Records Bulletin, No. 48. New York: Educational Records Bureau, 1947. Pp. xii+66.

Report of an expert committee on the work of psychologists and psychiatrists in the services. Privy Council Office. London: His Majesty's Stationery Office, 1947. Pp. 95. 2 shillings.

Technical data digest. Air Technical Intelligence. U. S. Army Air Forces, Headquarters Air Matériel Command, Wright Field, Dayton, Ohio. Vol. 12, Nos. 1-4. Pp. 41, 53, 65, 49.

INDEX OF SUBJECTS

- Absolute pitch, review of literature on, 249
- Activity, of animals, spontaneous, 293
- Additional tables of probability of "runs"
 - of correct responses in learning and problem solving, 276
- Animals, spontaneous activity of, 293
- Attitude methodology, opinion—, response to Crespi and Conrad, 171
- Army Air Forces, operational fatigue in, 34
- Child psychology, manual of, 162
- Comment on
 - the validity of personality questionnaires, 80
 - discussion of, 83
- Convulsive behavior in the rat, 201
- Correct responses, in learning, probability of, 276
- Corrigenda, 288
- Counseling, psychotherapeutic,
 - present status of, 297
 - discussion of, 564
- Critical review of the literature on absolute pitch, 249
- Discussion of Heinlein's comment on the validity of personality questionnaires, 83
- Discussion of Snyder's "The present status of psychotherapeutic counseling," 564
- Discussion of some causes of operational fatigue in the AAF, 34
- Effect, history and present status of the law of, 489
- Effective use of manipulative tests in industry, 130
- Emotional adjustment, relation to intellectual function, 283
- Encyclopedia of psychology, 280
- Fatigue, operational, in AAF, 34
- Festinger's evaluation of scale analysis, 451
- The history and present status of the law of effect, 489
- Illumination standards for effective and easy seeing, 435
- Industry, manipulative tests in, 130
- Intellectual function, relation of emotional adjustment to, 274
- Law of effect, history and present status of, 489
- Leadership studies, with reference to military problems, 54
 - note on, 466
- Manipulative tests, in industry, 130
- Manual of child psychology, 162
- Masking of speech, 105
- Mental traits, relative variability of, 1
- Military problems, review of leadership studies, 54
- Note on a review of leadership studies with particular reference to military problems, 466
- On Festinger's evaluation of scale analysis, 451
- Operational fatigue, in Army Air Forces, 34
- Opinion—attitude methodology, response to Crespi and Conrad on appraisal of, 771
- Personality questionnaires
 - validity of, 80
 - discussion of, 83
- Pitch, absolute, review of literature, 249
- Present status of psychotherapeutic counseling, 297
 - discussion of, 564
- Probability, of runs of correct responses in learning, 276
- Psychotherapeutic counseling
 - present status of, 297
 - discussion of, 564
- Qualitative data, treatment of by scale analysis, 149
 - Festinger's evaluation of scale analysis, 451
- Rat, convulsive behavior in, 201
- Relation of emotional adjustment to intellectual function—a note, 283
- Relative variability, of mental traits, 1
- Response to Crespi's rejoinder and Conrad's reply to appraisal of opinion—attitude methodology, 171
- Review of leadership studies with particular reference to military problems, 54
 - note on, 466
- Review of the literature on absolute pitch, 249
- Sampling in the revision of the Stanford-Binet scale, 413
- Sampling, selective, in psychological research, 267
- Scale analysis
 - treatment of qualitative data by, 149
 - on Festinger's evaluation of, 451
- Seeing, illumination standards for, 435

- Selective sampling in psychological re-
search, 267
- Speech, masking of, 105
- Spontaneous activity of animals, 293
- Standards, illumination, for effective and
easy seeing, 435
- Stanford-Binet scale, sampling in revision
of, 413
- Tests, manipulative, in industry, 130
- Treatment of qualitative data by scale
analysis, 149
on Festinger's evaluation of, 451
- Validity, of personality questionnaires, 80
discussion of, 83
- Variability, relative, laws of, 1

INDEX OF AUTHORS

ORIGINAL CONTRIBUTIONS, SHORT ARTICLES, SPECIAL REVIEWS, REPORTS, NOTES

Baier, D. E., 466
Barker, R. G., 162
Ellis, A., 83
Ellis, R. S., 1
Festinger, L., 149
Finger, F. W., 201
Grant, D. A., 276
Guttman, L., 451
Harris, R. E., 283
Heinlein, C. P., 80
Jenkins, W. O., 54
Johnson, H. M., 288
Kahn, L. A., 34

Lawshe, C. H., Jr., 130
Long, W. F., 130
Marks, E. S., 267, 413
McNemar, Q., 171
Miller, G. A., 105
Moreno, J. L., 564
Neu, D. M., 249
Postman, L., 489
Reed, J. D., 393
Snyder, W. U., 297
Thompson, C. W., 283
Tinker, M. A., 435
Wallace, S. R., 280

BOOKS REVIEWED

Abrahamson, D., 389
Angell, W., 185
Arthur, G., 572
Barker, R. G., 479
Baxter, E. D., 196
Beaumont, H., 97
Benedek, T., 179
Black, I. S., 197
Blair, G. M., 193
Blankenship, A. B., 391
Blumenfield, W., 190
Bridges, J. W., 475
Carp, B., 90
Casey, R. D., 94
de Gruchy, C., 582
Delgado, H., 586
Deutsch, A., 194
Dewey, J., 95
Dunlap, K., 195, 289
Edwards, A. E., 183
Fales, W., 188
Ford, M., 579
Friedman, B. B., 290
Gonick, M. R., 479
Gottschalk, L., 185
Gray, J. S., 476
Hall, V. E., 477
Harriman, P. L., 101
Hewitt, L. E., 180
Hopf, H. A., 191
Horsley, J. S., 570
Huntington, E., 92
Jenkins, R. L., 180
Johnson, W., 89
Kaback, G. R., 581
Kelley, D. M., 483
Kingsley, H. L., 292
Kluckhohn, C., 185
Landis, C., 568
Lasswell, H. D., 94
Lawton, G., 582

Leeper, R., 482
Lewis, C., 480
Lillie, R. S., 187
Link, H. C., 191, 587
Lowrey, L. G., 571
Luck, J. M., 477
Maier, N. R. F., 99
Mason, C. W., 192
Masserman, J. H., 87
Maurer, K. M., 578
Mays, A. B., 292
Michotte, A., 585
Minor, R. W., 392
Moncrieff, R. W., 295
Morgan, J. J. B., 90
Munn, N. L., 468
Radke, M. J., 576
Rand, W., 96
Roe, A., 387
Rogers, C. R., 88
Sargent, W. E., 189
Shakow, D., 388
Shroeder, E. M., 182
Smith, B. L., 94
Stern, E. M., 91
Sweeny, M. E., 96
Symonds, P. V., 177
Thornton, N., 389
Thorpe, L. P., 181
Townsend, A., 294
Traxler, A. E., 294
Vincent, E. L., 96
Wallen, J. L., 88
Witmer, H. C., 573
Wolff, W., 575
Wright, B. A., 479
Ammons, R. B., 95
Baldwin, A. L., 391, 576
Bayley, N., 577
Berg, I. A., 90
Bills, A. G., 477

BOOK REVIEWERS

- Britt, S. H., 391
 Brownell, W. A., 295
 Challman, R. C., 388
 Conrad, H. S., 585
 Cooley, N., 192
 Darrow, C. W., 392
 Dennis, W., 481
 Dillinger, C. M., 91
 Dollard, J., 569
 Duffy, E., 291
 Fiedler, M. F., 198
 Fischer, R. P., 188
 Gilmer, B. v. H., 296
 Goodenough, E. L., 575
 Graham, J. L., 292
 Guest, L. P., 101
 Guthrie, E. R., 88
 Harrell, T. W., 99
 Harris, D. B., 94
 Hartmann, G. W., 189
 Hartstein, J. I., 194
 Haslerud, G. M., 475
 Heider, F., 586
 Helson, H., 474
 Hertzman, M., 582
 Hilgard, E. R., 588
 Holzberg, J. D., 195
 Humphreys, L. G., 190
 Hunt, W. A., 89
 Jenkins, W. O., 193
 Jenness, A. F., 187
 Landis, C., 181
 Ligon, E. M., 290
 Lynch, C., 183
 Marzolf, S. S., 92
 Maurer, K. M., 182
 McKinney, F., 179
 McNemar, Q., 579
 Meltzer, H., 572
 Morgan, C. T., 479
 Pattie, F. R., 389
 Peterson, G. M., 102
 Porter, J. M., Jr., 294
 Richards, T. W., 180
 Sappenfield, B. R., 196, 483
 Seashore, R. H., 192
 Shipley, W. C., 389
 Spoerl, H. D., 191, 587
 Stagner, R., 480
 Stone, L. J., 581
 Switzer, St. C. A., 90
 Terman, L. M., 484
 Thompson, A. S., 197
 Thompson, G. G., 97
 Wells, W. R., 96
 Youtz, A. C., 573
 Zubin, J., 185, 571

✓



